

```
Query Match: 18.17% Indels: 24
DB: 9 Gaps: 5
US-09-921-992-50 (1-372) x US-09-921-992-22 (1-670)
QY 133 ArgIleGlyValAsnAlaGlySerLeuGlu-LysAspLeuGlnGluLysTyrGlyGluPr 152
DB 15 AGGATCGCGGTTAACCATGGTCTCTCTCCGAAAGAATGC-----TTTACCTACTTTT 68
QY 152 oThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAs 172
DB 69 TACCCCGAGGGATGGTGAATTCATCAAAATTTGTGAGTCCTTAGA 128
QY 172 nPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSerTy 192
DB 129 TTTCCGCAACCTAGTCTGTTTCCATGAAGCGTCCCGGTACCGGTATATGTTGCCCGCTA 188
QY 192 rArgLeuLeuAlaLysGlnIleAspGln-----ProLeuHisLeuGlyI 207
DB 189 TCGCTCATGGTGAACAGTATGACGAGTTGGGCATGGATTATCCCTCCATCTAGGGT 248
QY 207 eThrGluAlaGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 227
DB 249 TACCGAAGCGGGATGGGAATATGCGCGCATTAATCCACCGTGGCATTCGCCACCT 308
QY 227 uLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 247
DB 309 TTTAGCTGATGGCATTTGGCGATATATCCGGTATCCCTCACCAGAGCCCCGAAAGA 368
QY 247 uIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPh 267
DB 369 AATTCCCGTTGTACAGCATTTCTCAGGCGTGGTTCGCGAAACCATGTTGGAATA 428
QY 267 eIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 287
DB 429 TGTGGCTGTCTCTCTGTCGCGCACGTTGTTCAAC-----TT 467
QY 287 uGluGlnArgLeuGluAspIle-----IleThrProMetAsp-ValS 301
DB 468 GGAAGACGTGTACATGAAGTCCGAGATGCCACTAAACATCTAACCTGTTTAGCTTCG 527
QY 301 erIleLeGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 321
DB 528 NCGTCATGCTGTATTGTCAATGGCCCGGTGCAATGGCCGATGCGCATATGGCTATG 587
QY 321 hrGly 322
DB 588 TGGGT 592
RESULT 15
US-09-921-992-23
; Sequence 23, Application US/09921992
; Patent No. US20020069426A1
; GENERAL INFORMATION:
; APPLICANT: Boronati, Albert;
; APPLICANT: Campos, Narciso;
; APPLICANT: Rodriguez-Concepcion, Manuel;
; APPLICANT: Rohmer, Michel;
; APPLICANT: Seeman, Myriam;
; APPLICANT: Valentin, Henry E.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; APPLICANT: Venkatesh, Mylavaraopu
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
; FILE REFERENCE: 16516.107/35-21(51897)US
; CURRENT FILING DATE: 2001-08-06
; PRIOR FILING DATE: 2000-08-06
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 23
; LENGTH: 596
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1..596)
; OTHER INFORMATION: unsure at all n locations
US-09-921-992-23
Alignment Scores: 1.71e-30 Length: 596
Pred. No.: 326.00 Matches: 78
Score: 58.62% Conservatives: 41
Best Similarity: 38.42% Mismatches: 62
Best Local Similarity: 17.47% Indels: 22
Query Match: 9 Gaps: 4
US-09-921-992-50 (1-372) x US-09-921-992-23 (1-596)
QY 133 ArgIleGlyValAsnAlaGlySerLeuGlu-LysAspLeuGlnGluLysTyrGlyGluPr 152
DB 15 AGGATCGCGGTTAACCATGGTCTCTCTCCGAAAGAATGC-----TTTACCTACTTTT 68
QY 152 oThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAs 172
DB 69 TACCCCGAGGGATGGTGAATTCATCAAAATTTGTGAGTCCTTAGA 128
QY 172 nPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSerTy 192
DB 129 TTTCCGCAACCTAGTCTGTTTCCATGAAGCGTCCCGGTACCGGTATATGTTGCCCGCTA 188
QY 192 rArgLeuLeuAlaLysGlnIleAspGln-----ProLeuHisLeuGlyI 207
DB 189 TCGCTCATGGTGAACAGTATGACGAGTTGGGCATGGATTATCCCTCCATCTAGGGT 248
QY 207 eThrGluAlaGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 227
DB 249 TACCGAAGCGGGATGGGAATATGCGCGCATTAATCCACCGTGGCATTCGCCACCT 308
QY 227 uLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 247
DB 309 TTTAGCTGATGGCATTTGGCGATATATCCGGTATCCCTCACCAGAGCCCCGAAAGA 368
QY 247 uIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPh 267
DB 369 AATTCCCGTTGTACAGCATTTCTCAGGCGTGGTTCGCGAAACCATGTTGGAATA 428
QY 267 eIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 287
DB 429 TGTGGCTGTCTCTCTGTCGCGCACGTTGTTCAAC-----TT 467
QY 287 uGluGlnArgLeuGluAspIle-----IleThrProMetAsp-ValS 301
DB 468 GGAAGACGTGTACATGAAGTCCGAGATGCCACTAAACATCTAACCTGTTTAGCTTCG 527
QY 301 erIleLeGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 321
DB 528 NCGTCATGCTGTATTGTCAATGGCCCGGTGCAATGGCCGATGCGCATATGGCTATG 587
QY 321 hrGly 322
DB 588 TGGGT 592
Search completed: November 23, 2003, 21:59:12
Job time : 4946 secs
```

THIS PAGE BLANK (USPTO)

QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgMetValValAsp 123
Db 700 AAGATCGTGTCAACCCAGGAATTTGCGGACAGCGGCCCGCCAGTTTGAGACGATGAT 759
QY 124 CysAlaArgAspLys----- 128
Db 760 TATACAGAAGATGAATATCATCAGAAAGAACTCCAGCATATCGAGCAGGTCTTCACTCCTTTG 819
QY 129 -----AsnIleProIleArgIleGlyValAsnAlaGlySerLeu 141
Db 820 GTTGAGAAATGCAAAAAGTACGGAGAGCAATCGGTATTGGGCAAAATCATGGAAGTCTT 879
QY 142 GluLysAspLeuGlnGluLysTyrglyGluProThrProGlnAlaLeuLeuGluSerAla 161
Db 880 TCTGACCGTATCATGACGTATTACGGGAT--TCTCCCGAGGAATGGTTGAATCTGCG 936
QY 162 MetArgHisValAspHisLeuAspArgLeuAsnPhaAspGlnPhaLysValSerValLys 181
Db 937 TTTGAGTTTGCAGAAATATGTCGGAATATTAGACTATCACAACTTTGTTTCTCAATGAAA 996
QY 182 AlaSerAspValPheLeuAlaValGluSerTyrglyGluLeuAlaLysGlnIle----- 199
Db 997 GCGAGCAACCAATGATCATGGTCCAGGCGTACGTTTACTTGTGCTGAGATGTATGTT 1056
QY 200 -----AspGlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGly 216
Db 1057 CATGGATGGATATCTTTGCAATTTGGAGTTACTGAGCAGAGAGCGGAGATGGA 1116
QY 217 AlaValLysSerAlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeu 236
Db 1117 CGGATGAAATCTCGATTGGAATTTGGACGCTTCTTCAGACGGGCTCGGTGACACAATA 1176
QY 237 ArgValSerLeuAlaAlaAppProValGluGluIle 248
Db 1177 AGAGTTTCACTGACGGAGCCACCAGAGAGGAGATA 1212

RESULT 13

US-10-066-543-1184
; Sequence 1184, Application US/10066543
; Publication No. US20030087818A1

GENERAL INFORMATION:

; APPLICANT: Jiang, Yugu
; APPLICANT: Xu, Jiangchun
; APPLICANT: Indrias, Carol Yoseph
; APPLICANT: Lodes, Michael J.
; APPLICANT: Secrist, Heather
; APPLICANT: Carter, Darrick
; APPLICANT: Fanger, Gary R.
; APPLICANT: Smith, Carole L.
; APPLICANT: Durham, Margarita
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.563
; CURRENT APPLICATION NUMBER: US/10/066,543
; CURRENT FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 3417
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1184
; LENGTH: 706
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 2, 645, 659
; OTHER INFORMATION: n = A,T,C or G

US-10-066-543-1184

Alignment Scores: 5.99e-46 Length: 706
Pred. No.: 452.00 Matches: 89
Score: 73.29% Conservative: 29
Percent Similarity:

Best Local Similarity: 55.28% Mismatches: 43
Query Match: 24.22% Indels: 0
Db: 14 Gaps: 0

US-09-921-992-50 (1-372) x US-10-066-543-1184 (1-706)

QY 10 ArgLysSerThrArgIleTyrglyAsnValProIleGlyAspGlyAlaProIleAla 29
Db 218 AAAGAGAGTCGAGAAATACAAATCGGATGTTAAGATTGGGAAAGAAATCCCGTTGTG 277
QY 30 ValGlnSerMetThrAsnThrArgThrAspValGluAlaThrValAsnGlnIleLys 49
Db 278 ATTCAATCGATGATTAAACACGGAATCTGCGGATGTAGAACGCGAGTGTACAGCAAAATTTTG 337
QY 50 AlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAlaAla 69
Db 338 GATTTTGAAGAAGCGCGTGTGAATTTGTTCTGTATGACGATCAATACAAAGAGCGCGCT 397
QY 70 GluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAspIleHisPhe 89
Db 398 ATGGCAATTTCCCGCGATTTAAAGAGAGTTTCATATTTCTTTGTTAGCGGATATTCATTT 457
QY 90 AspTyrglyIleAlaLeuLysValAlaGluTyrglyValAspCysLeuArgIleAsnPro 109
Db 458 GATTATCGACTTGCCTTATTGCGGATAGAACAGGGAATTCACAAATTTACGGATTAATCCC 517
QY 110 GlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLysAsn 129
Db 518 GGAATATATCGTTCGAGAAATATTCGTTTGGTTCGGAAGCGGCAAAAGGAAGAGA 577
QY 130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrgly 149
Db 578 ATTCGATTCGATTGAGTCAATGCAGGTTCTTTGGAAAAACATATCTTGGAAAAATAT 637
QY 150 GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAsp 169
Db 638 GGAGCGGNAACGCGAGATGCTNTGTCAAAGCGCTATGATCATGTAAAACTGTTGGAA 697
QY 170 Arg 170
Db 698 CAA 700

RESULT 14

US-09-921-992-22
; Sequence 22, Application US/09921992
; Patent No. US20020069428A1

GENERAL INFORMATION:

; APPLICANT: Boronot, Albert;
; APPLICANT: Campos, Narciso;
; APPLICANT: Rodriguez-Concepcion, Manuel;
; APPLICANT: Rohmer, Michel;
; APPLICANT: Seeman, Myriam;
; APPLICANT: Valentin, Henry E.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; APPLICANT: Venkatramesh, Mylavara
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
; FILE REFERENCE: 16516.107/35-21(51897)US
; CURRENT APPLICATION NUMBER: US/09/921,992
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/223,483
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 22
; LENGTH: 670
; TYPE: DNA
; ORGANISM: Zea mays
; OTHER INFORMATION: n = A,T,C or G

US-09-921-992-22

Alignment Scores: 5.06e-32 Length: 670
Pred. No.: 339.00 Matches: 81
Score: 59.81% Conservative: 44
Percent Similarity: 38.76% Mismatches: 60
Best Local Similarity:


```
Db 303 TTCCATTATAAATTCGCTCTC---ATTGCGCGCTCAAAAGCGTGGATCGCATCAGGATTAAC 359
Qy 109 ProGlyAsnIleGlyAsnGluArgIleArgMetValValAspCysAlaArgAspLys 128
Db 360 CCGGAAACATCGCTCTTAAGAGAAGATCAAAAGCGTGGTGTATCTGTTAAAGAAAAA 419
Qy 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
Db 420 AACATTCTTAAGATTGGCGTGAATGCTGGGAGTTTAGAAAGCAGTTTGTATCAAAAA 479
Qy 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
Db 480 TACGGA---CCCAACCCAAAGCATCGTAGAAAGCGCTTTGTATTAACGCCAAACTTTTA 536
Qy 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db 537 GAAGATTGGATTTTACCAATTTTAGATTCTTTAAAGCGGAGCGATGATTCACACC 596
Qy 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db 597 ATAGAGCTTACAGGATGCTTCGCCCTCTGTGATCTATCTTTCATTGGGGTTACG 656
Qy 209 GluAlaGlyCysAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
Db 657 GAGCGGGGAATCTTTTAGCTCCAGTATCAATCCGCTATGCTTTAGGGGGCTTTTA 716
Qy 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
Db 717 ATGAGGGCAATGGGATGAGTGGCGGTATCTCATCAGGGGAATTAGAAATGAATC 776
Qy 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
Db 777 AAGTGGCCAGACCAATTTTACCCATAGCGCGGCTTGAAGAGGGAATTAATTGGATT 836
Qy 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
Db 837 TCTTGCCCACTTCGCGGCGCATTAAGCCCAATTTAGTGGATATGCGATCAAGGTAGA 896
Qy 289 GlnArgLeuGluAspIleThrProMetAspValSerIleIleGlyCysValValAsn 308
Db 897 AAACGCTTAAGCCACATCAAAACCCCTTTAGACATTAGCGTGATGGGTTCGGTGAAT 956
Qy 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly 328
Db 957 GCTTTGGGTGAAGCAAGCATGCAGATCGCATCGCTTTTGGGAATCGCAGCGGTTG 1016
Qy 329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeu 348
Db 1017 ATCATTAAAGAGGGTAAAGTCATTACAAACTGGCTGAAAAGGATTATTATTGAAACTTTT 1076
Qy 349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
Db 1077 GTGATAGAGTGGAAATTTAGCTTAAGAAAGAGAAAGAAAGT 1118
```

RESULT 8

```
US-10-156-761-1636
; Sequence 1636, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: HATTORI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; FILE OF INVENTION: NOVEL POLYNUCLEOTIDES
; TITLE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
```

```
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1636
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1155)
US-10-156-761-1636
```

```
Alignment Scores:
Pred. No.: 5,936-87
Score: 786.50
Percent Similarity: 63.69%
Best Local Similarity: 44.44%
Query Match: 42.15%
DB: 14
Matches: 1155
Conservative: 71
Mismatches: 121
Indels: 13
Gaps: 4
```

US-09-921-992-50 (1-372) x US-10-156-761-1636 (1-1155)

```
Qy 6 Profile---GlnArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAsp 24
Db 43 CCGATCGCGCAAGCGGTGTCTCCCGGAGATCCAGTCCGACCGCTGGCGGTGGCGGC 102
Qy 25 GlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThr 44
Db 103 ACGGCCCCGTGTCGTGTCAGTCGATGACGACGACGCGTACGTCCGACATCGCGCACG 162
Qy 45 ValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValPro 64
Db 163 CTCAGCAGATCGCGGAATCTACCGCGTCGCGCTGCAGATCGTCGCGCTCGCTCCGCC 222
Qy 65 ThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsnValProLeuVal 84
Db 223 ACGCAGGACGACGCGGACGCCCTCGCGGTATCGCCGCAAGTCGAGATCCCGTCTGTC 282
Qy 85 AlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCys 104
Db 283 GCGGACATCCACTCCAGCGCAAGTACGTGTTCCGCGCATCGAGCGGGCTGCGCGCGC 342
Qy 105 LeuArgIleAsnProGlyAsnIleGlyAsn---GluGluArgIleArgMetValValAsp 123
Db 343 GTCCGGGTGAACCCGGGCAACATCAAGCAGTTCGACACAAGTCAAGGAGATCGCGCGC 402
Qy 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143
Db 403 GCGCCCAAGGAGACGCGCAGCGCATCCGATCGCGGTCAACGCGGCTCGCTCGACCGG 462
Qy 144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163
Db 463 CGGCTGCTGGAGAAGTACGCAAGGCCACCCCGAGGCGCTCGTCGAGTCGGCTGTGG 522
Qy 164 HisValAspHisLeuAspArgIleAsnPheAspGlnPheLysValSerValLysAlaSer 183
Db 523 GAGGCGTCTCTTCGAGGACGACACTTCCGCGGACATCAAGATCTCGGTCAAGCAAC 582
Qy 184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203
Db 583 GACCCGGTGTATGTCATGTCACCGGTACCGCAGCTGGCGCGCGGAGTCCGACTACCGCTG 642
Qy 204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyValaValLysSerAlaIleGly 223
Db 643 CACCTCGGCTGACCGAGCGCGCCCGCTTCCAGGCGACGATCAAGTCGCGCGCTGCC 702
Qy 224 LeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243
Db 703 TTCGGCGCGCTGCTCAGCGAGGCGATCGCGCACACGATCCGCTCTCGTTCGACGCGCC 762
Qy 244 ProValGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263
Db 763 CCGTTCGAGGAGATCAAGTTCGCGATCCAGATCTCTGGAGTCTGGGCTCTCAGCGCGG 822
```

; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:	
Pred. No.:	1,11e-83
Score:	808.00
Percent Similarity:	64.11%
Best Local Similarity:	44.93%
Query Match:	43.30%
DB:	14
Length:	9025608
Matches:	164
Conservative:	70
Mismatches:	119
Indels:	12
Gaps:	3

US-09-921-992-50 (1-372) x US-10-156-761-1 (1-9025608)

Qy	8	GlnArgArgIysSerThrArgIleTyrValIcIyAsnValProIleGlyAspGlyAlaPro	27
Db	3141566	GAGCGCCGGAAGAGCCGCGAGATCCAGGTCCGAACCGCTGGCGCGGAGACGCACCC	3141507
Qy	28	IleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGln	47
Db	3141506	GTCTCGGTGCAGTCCGATGAGCAGCCGCTACGTCGACATCCGCGCACCGCTCCACGAG	3141447
Qy	48	IleIysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAsp	67
Db	3141446	ATCGCCAGACTCACGGCGTCCGGCTGCCAGATCGTCTGCTGGCTCCCCACGCGAGAC	3141387
Qy	68	AlaAlaGluAlaPheIysLeuIleLysGlnValAsnValProLeuValAlaAspIle	87
Db	3141386	GAGCGGAGCCCTCGCGGTATCGCCCGCAAGTCGCGAGATCCCGGTGATCGCGGACATC	3141327
Qy	88	HisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIle	107
Db	3141326	CACCTTCAGCCGGAAGTAGTACGTTCGCGCGCATCGAGCGGGCTCGCGCGGTCTCCGGTG	3141267
Qy	108	AsnProGlyAsnIleGlyAsn--GluGluArgIleArgMetValValAspCysAlaArg	126
Db	3141266	AACCGGGCAACATCAAGCAGTTCGACGACAGAGTCGAAGGAGATCCCAAGCGCGGGAAG	3141207
Qy	127	AspIysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGln	146
Db	3141206	GAGCACGGCAGCGCATCCGATCGCGCTCAACGGGGCTCGCTCGACCGCGGTGCTG	3141147
Qy	147	GluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp	166
Db	3141146	GAGAAGTACGCGCAAGGCACCCCGAGGCGCTCGTCAGTTCGGCTGTGGAGGGCGTCG	3141087
Qy	167	HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValIysAlaSerAspValPhe	186
Db	3141086	CTCTTCGAGGAGCAGCATCTCCGGACATCAAGATCTCGGTCAAGCAACACGACCCGGTG	3141027
Qy	187	LeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGly	206
Db	3141026	GTATGGTCAACGGGTACCGCCAGTTCGGCGGCGAGTCGACATCCCGCTGCACTTCGGC	3140967
Qy	207	IleThrGluAlaGlyGlyAlaArgSerGlyValAlaLysSerAlaIleGlyLeuGlyLeu	226
Db	3140966	GTGACCGAGCGCGCCCGCTTCAGGGGCACGATCAAGTCGCGCTCGCTCGCGCGCGG	3140907
Qy	227	LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu	246
Db	3140906	CTGCTCAGCGAGGGCATCGCGACACGATCCGCGTCTCTCGTGAAGCGCGCCCGCTCGAG	3140847
Qy	247	GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn	266
Db	3140846	GAGATCAAGTCCGCGATCCAGATCCTGGAGTCTGTAACCTCGCGCAGCGCGCGCTGGAG	3140787
Qy	267	PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla	286
Db	3140786	ATCGTCTCTCCGCTCTCGCGAGCGGCCGACGCTCTACAAGTGGCCGCGAGGAG	3140727
Qy	287	LeuGluClnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysVal	306
Db	3140726	GTACGCGCGGTCTGGAGGCATGAGAGGTCCCTCGCGGTCTCGGCTATGGCTCGCTG	3140667

Db 924 GGTCCGGAGAGCGCGAAGCCGATATCCGGATCCGGCGGCGACGCGG 974

RESULT 5

US-10-156-761-2548

; Sequence 2548, Application US/10156761

; Publication No. US20030119018A1

; GENERAL INFORMATION:

; APPLICANT: OMURA, SATOSHI

; APPLICANT: IKEDA, HARUO

; APPLICANT: ISHIKAWA, JUN

; APPLICANT: HORIKAWA, HIROSHI

; APPLICANT: SHIBA, TADAYOSHI

; APPLICANT: SAKAKI, YOSHIYUKI

; APPLICANT: HATTORI, MASAHIRA

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

; FILE REFERENCE: 249-262

; CURRENT APPLICATION NUMBER: US/10/156,761

; PRIOR FILING DATE: 2002-05-29

; PRIOR APPLICATION NUMBER: JP 2001-204089

; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697

; PRIOR FILING DATE: 2001-08-02

; NUMBER OF SEQ ID NOS: 15109

; SEQ ID NO 2548

; LENGTH: 1155

; TYPE: DNA

; ORGANISM: Streptomyces avermitilis

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)..(1155)

US-10-156-761-2548

Alignment Scores:

Pred. No.: 1,31e-89 Length: 1155
Score: 808.00 Matches: 164
Percent Similarity: 64.11% Conservative: 70
Best Local Similarity: 44.93% Mismatches: 119
Query Match: 43.30% Indels: 12
DB: 14 Gaps: 3

US-09-921-992-50 (1-372) x US-10-156-761-2548 (1-1155)

QY 8 GlnArglySerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaPro 27
DB 49 GAGCGCGGAGAGCGCGGAGATCCAGGTCCGACCGTGGCGGCGGAGCGACCC 108
QY 28 IleAlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThrValAsnGln 47
DB 109 GTCTCGGTGCGATGACGAGCGCGTACGTCCGACATCGCGCGCGACGCTCCAGCAG 168
QY 48 IleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAsp 67
DB 169 ATCGCGAGCTCAGCGGTCCGCGTCCAGATCGTGGTGGCGTCCGCGCGAGGAC 228
QY 68 AlaAlaGluAlaPheLysLeuLysGlnValAsnValProLeuValAlaAspIle 87
DB 229 GAGCGGAGCCCTCGCGGTCCATCCCGCGGAGTCCGAGATCCCGTGTGCGGACATC 288
QY 88 HisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIle 107
DB 289 CACTTCCAGCGAAGTACGTGTTCCGCGCGATCGAGCGGCGTCCGCGGCGTCCGCGGTG 348
QY 108 AsnProGlyAsnIleGlyAsn---GluGluArgIleArgMetValValAspCysAlaArg 126
DB 349 AACCCGGGCAACATCAAGCAGTTCGACGACAGGTCAAGGAGATCGCCAAAGCGCGGAG 408
QY 127 AspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGln 146
DB 409 GAGCAGCGGCGCGATCCGATCCGCGGTCAACCGCGGTCTCGTCCGCGCGGTCTGTG 468
QY 147 GluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp 166
DB 469 GAGAAGTACGCAAGGACCCCGGCGGTCTGTGAGTCCGCGGTCTGTGGGAGGCGTGTG 528


```
Db 314632 GAAATCAATTTCTTACGCAAAAGATGAAATATTCCTAATTCGTATGTTAATGCTGGA 314691
Qy 140 SerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGlu 159
Db 314692 TCITTAGAAAAGATATATTAATAAATAATAAATACCTACTCCAGATGCTTAGTAGAA 314751
Qy 160 SerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSer 179
Db 314752 TCAGCTATGAGGCATATTGAATCTTTGATGCTTTAAATTTTAAATTTAAATTTAGT 314811
Qy 180 ValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnLe 199
Db 314812 GTTAAAGCGTCTGATGATTTTAGCTATTGAATCATATCGCATGTTAGGAAAAAATTT 314871
Qy 200 AspGlnProLeuHisLeuGlyLeuThrGluAlaGlyAlaArgSerGlyAlaValLys 219
Db 314872 ACACAACCTTTGCATATGCAATACTGAATCCGGTGGTTAAGGAATGGAACAGTTAA 314931
Qy 220 SerAlaIleGlyLeuLeuLeuLeuSerGluGlyLeuGlyIleGlyAspThrLeuArgValSer 239
Db 314932 TCATCTATAGGTATGCTTTTATTATTATTAGAAAGGCATTGGAGATACATACAGATTCA 314991
Qy 240 LeuAlaAlaAspProValGluGluLysValGlyPheAspIleLeuLysSerLeuArg 259
Db 314992 TTAGCGGCACATCCAACTGAAGAAGTAAAGTAGTTATGACATTTTAAAGTTTATCT 315051
Qy 260 IleArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAsp 279
Db 315052 TTAAAGCAAGAGGTATTAATTTATGCTTGTCTTCTAGTGTCTAGCAAGAAATTTGAT 315111
Qy 280 ValIleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAsp 299
Db 315112 GTAATTAATACAGTAAATCAACTAGAAAAAATCTAGAGATATCTCGACTCCCATAGT 315171
Qy 300 ValSerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGly 319
Db 315172 GTATCAATATTGTTGCGTGTGTTAATGAATAGGTGAATCAATAAGCACTTTAGGT 315231
Qy 320 ValThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArg 338
Db 315232 CTAGCAGGAAGTCAATGAAGAAAGTGTCATTTTATGAAGCGGAGTAAAGCAAAAGAAAA 315291
Qy 339 LeuAspAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGln 358
Db 315292 ATAAAAACGAGAAATATTAGAAAAATGGAATTAATTAATTCGAAAAAATAGACAAA 315351
Qy 359 LeuAspGluAlaArgArg 364
Db 315352 TTAATAATAATTCAAAAAAA 315369

RESULT 4
US-09-974-300-1692
; Sequence 1692, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1692
; LENGTH: 1083
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1692
```

```
Alignment Scores: 2-21e-91 Length: 1083
Pred. No.: 822.00 Matches: 162
Score: 68.45% Conservativeness: 55
Best Local Similarity: 51.10% Mismatches: 100
Query Match: 44.05% Indels: 0
DB: 10 Gaps: 0

US-09-921-992-50 (1-372) x US-09-974-300-1692 (1-1083)

Qy 9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
Db 24 CGTACAAAACCGCTCCCGTTAAAGTGGAGCCCTTAAACAATAGCGCGCAATACGAAGTC 83
Qy 29 AlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThrValAsnGlnIle 48
Db 84 GTCAATTCGAAGCATGACAAACGAAACACATGATGCTTGAAGCAACCGTCGCCGAAATC 143
Qy 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
Db 144 AACAGACTCGCGGAAGCAGGATGTCAAATCGTCCGCGTCCCTGCTCGATGAACGGGCT 203
Qy 69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAspIleHis 88
Db 204 GCGAGCGCATTCACAGATCAAAAAGCGGATATCCATCCCTCTTGTGCGTGAATTCAT 263
Qy 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
Db 264 TTCAACTATAATTTGGCATTTAAAGCATCGAAGCGGAGCGGCAATAAAATCCGATCAAT 323
Qy 109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
Db 324 CCGGTAAATCATCGCGCGCGGAAAGGTGAAGCGGTGCTCAACGCGAGCGAAGAAAG 383
Qy 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
Db 384 GGCATTCGATCCGATCGGATCGGCTCAATCGAGCTCTCTGGAAAAACGAATCTTCGAGAG 443
Qy 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
Db 444 TAGCGCTATCCGACAGCAGCAGCGCATGTCGAAAGCGCCCTGCACCAACATTAAAAATCTT 503
Qy 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db 504 GAGGATCTCGATTTCCACGATATCATCTGTCAGCATGAAGCGGTCTGTGTTAAACCTGGC 563
Qy 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db 564 ATTGAGGCATATGAAAAAGCGGCTAAAGCTTTTCGATTTATCCGCTTCATTTAGGCATACC 623
Qy 209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
Db 624 GAATCGGAACCGCTGTTTTCGCGTACAGTGAAGAAAGCGCGCGGCTCTCGCGCGCATCTT 683
Qy 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluLe 248
Db 684 TCAAAAGGAATCGGCAATACATTCGGATTTCTTTAAAGCGCGGACCCCGGTGCAAGAGTA 743
Qy 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
Db 744 AAAGTCGCCGAGGAGCTCTGAAATCTTCGGGCTCGCTTCCAATCGGCAACATTGATT 803
Qy 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
Db 804 TCCTGCCGAGCTTCGGCGCGGATCGAAATCGATTTGATCTGCAATTCGCAATGAATTCGAA 863
Qy 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
Db 864 GATTACATTCGCGAAAAATCAAGCGCGCGGATCAAGGTTGCGGTTCTCGGCTGTGCGGTCAC 923
Qy 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLys 325
```



```
Alignment Scores:
Pred. No.: 5,01e-183 Length: 1830121
Score: 1603.50 Matches: 320
Percent Similarity: 93.66% Conservative: 20
Best Local Similarity: 88.15% Mismatches: 22
Query Match: 85.93% Indels: 1
DB: 14 Gaps: 1

US-09-921-992-50 (1-372) x US-10-329-960-1 (1-1830121)

QY 4 GlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23
|||
|||
|||
Db 390972 CAGCCAACTATTAAAGCGTCGTGAATCGACAAAATTTATGTGGAAATGTACCAATTTGGT 391031
QY 24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAla 43
|||
|||
|||
Db 391032 GGGATGCGGCTATTGCGTGCATCAATGACAAATCTCGCACCACTGATGTGGAAGCG 391091
QY 44 ThrValAsnGlnIleLysAlaLeuGluAlaValGlyAlaAspIleValArgValSerVal 63
|||
|||
|||
Db 391092 ACAGTTGCTCAATTAATCATTTAGAACGTTGTTGGTCAGATATTTGTTGTTGTTCTGTT 391151
QY 64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeu 83
|||
|||
|||
Db 391152 CCAACAATGATGTCGCGGAAGCATTTAAACAAATTAACAAAGTGAATGTTCCGCTC 391211
QY 84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103
|||
|||
|||
Db 391212 GTACGAGATATTCATTTGCACTATCGTATCGCTTAAAGTCGAGAATATGAGTGGAT 391271
QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123
|||
|||
|||
Db 391272 TGTTTACGTATCAATCTGCGCAACATTTGTCGTAAGATCGCTCGCTGCTGTTGAT 391331
QY 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143
|||
|||
|||
Db 391332 TGTGCGGAGACAAAATATTCGATTCGTATTGTTGTAATGCAGGCTCTTTAGAAAAA 391391
QY 144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGluSerAlaMetArg 163
|||
|||
|||
Db 391392 GATTTGCAAGAAAATATGCGCAACCAACCGCAGAACCTTTGTAGAACCTCGCATCGCGT 391451
QY 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183
|||
|||
|||
Db 391452 CATGTAGAAATCTTAGATCGCTTAATCTTCACTTCACTTAAAGTACGCTGTAAGCGCTCC 391511
QY 184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203
|||
|||
|||
Db 391512 GATGTTATCTTAGCGGTTGAATCTTATCGTTTACTGGCTAAAGCAATTAACACAGCCTTTA 391571
QY 204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223
|||
|||
|||
Db 391572 CATTTAGGCATTCAGAACGAGTGGCGCACGGGCTGGTGCACTAAATCTGCAAGTGGGT 391631
QY 224 LeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243
|||
|||
|||
Db 391632 TTAGGAATGTTATAGCTGAGGCGATTTGGCGATACACTACGCTCTCTTTGGCGGCGAGAT 391691
QY 244 ProValGluGluIleLysValGlyPheAspIleLysSerLeuArgIleArgSerArg 263
|||
|||
|||
Db 391692 CCTGTAGAGGAATCAAAAGTCGGTTTGATATTTGAAATCTTTACGGATTCGTTCAAGA 391751
QY 264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283
|||
|||
|||
Db 391752 GGAATTAACCTTTATGCTTCCCAACCTGTTCTCGCAAGAAATTTGATGTAATCGGTACA 391811
QY 284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303
|||
|||
|||
Db 391812 GTAAATGCGCTAGAACCAACCCCTTGAAGATATTATTACCAATGATGATCTATTATC 391871
QY 304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323
|||
|||
|||

Db 391872 GGTGTGTGTAGTGAATGGTCTCGCGAGGCACTCGTCTCGCATCTCGCGCTAAACGGCGGT 391931
QY 324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAsn 342
|||
|||
|||
Db 391932 AACAAAAAAGCGGTTATTATCTTGACGGAGAACGCCAAAAAGAGCGTTTGTATAACGAA 391991
QY 343 AspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
|||
|||
|||
Db 391992 GATATAGTGAACCAATTAGAACAAAAATTCGTGCGAAAGTCGACAGACAGATCCAAAA 392051
QY 363 ArgArgIle 365
|||
|||
|||
Db 392052 AACAGAAAT 392060

RESULT 3
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEKI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
; US-09-790-988-1

Alignment Scores:
Pred. No.: 5,36e-144 Length: 640681
Score: 1282.00 Matches: 242
Percent Similarity: 83.88% Conservative: 65
Best Local Similarity: 66.12% Mismatches: 57
Query Match: 68.70% Indels: 2
DB: 10 Gaps: 2

US-09-921-992-50 (1-372) x US-09-790-988-1 (1-640681)

QY 1 MetHisAsnGlnAlaPro---IleGlnArgArgLysSerThrArgIleTyrValGlyAsn 19
|||
|||
|||
Db 314272 ATGAATAATAAGTGTAAATATTAATCAATAGAGAAAATCTGATGTTATTTATGTTGAAAA 314331
QY 20 ValProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThr 39
|||
|||
|||
Db 314332 GTCGCTATTGGCAATAATGCGCAATATCAGTTCAATCTATGACAAATACTCGTACTACT 314391
QY 40 AspValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleVal 59
|||
|||
|||
Db 314392 AATACTCTGAAACTATTATTAATCAATCTTAGAGTTACAAAAAGTAGGAGTAGATATTGTT 314451
QY 60 ArgValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnVal 79
|||
|||
|||
Db 314452 CGTATTTCTATACCAATTTAAAGCTGAGAAATCATTTCAAGAAAATAAAAAACAAACA 314511
QY 80 AsnValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGlu 99
|||
|||
|||
Db 314512 AATGTTCCATTTGTCAGATATACATTTTGTATGATACAGATTAGCTTTTACAGCTATATAAA 314571
QY 100 TyrGlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArg 119
|||
|||
|||
Db 314572 TATGTTGCGAGATTGTTTGAAGATTAATCTGGAATATTGGAATATAAAGAGAGTATCA 314631
QY 120 MetValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGly 139
|||
|||
|||
```

```
;/ NAME/KEY: misc_feature
;/ LOCATION: (44416)..(44416)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (44905)..(44905)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (44975)..(44975)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (45593)..(45593)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (45732)..(45732)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (47036)..(47036)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51334)..(51334)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51602)..(51602)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51786)..(51786)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (51805)..(51805)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (55369)..(55369)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (65309)..(65309)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (65313)..(65313)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (80024)..(80024)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (100091)..(100091)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (102696)..(102696)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (105121)..(105121)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (107248)..(107248)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
```

```
;/ LOCATION: (117136)..(117136)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (119750)..(119750)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (119924)..(119924)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (120038)..(120038)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (121344)..(121344)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (122167)..(122167)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (122336)..(122336)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (131340)..(131340)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (131360)..(131360)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (139910)..(139910)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (140398)..(140398)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (142750)..(142750)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (145058)..(145058)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (145171)..(145171)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (145942)..(145942)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (147197)..(147197)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (150841)..(150841)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (152500)..(152500)
;/ OTHER INFORMATION: n equals a, t, g or c
;/ FEATURE:
;/ NAME/KEY: misc_feature
;/ LOCATION: (152530)..(152530)
```

```
; NAME/KEY: CDS
; LOCATION: (1)...(1119)
US-09-921-992-3

Alignment Scores:
Pred. No.: 2,37e-220 Length: 1119
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 9 Gaps: 0

US-09-921-992-50 (1-372) x US-09-921-992-3 (1-1119)
QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
DB 1 ATGCATTAACGAGCTCCAAATTCACGCTAGAGAAATCAACACGATATTACGTTGGGAATGG 60
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
DB 61 CCGATTGGCGATGGTCTCCCATCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
DB 121 GTCGAAGCAACGGTCAATCAAAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGTCCGT 180
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheIleLysLeuIleLysGlnValAsn 80
DB 181 GTATCCGTACCGACGATGGACGGCGGAGAGCGCTTCAAACTCATCAACAGCAGGTTAAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
DB 241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTCGCGTGAAGTAGCGGAATAC 300
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
DB 301 GGGCTCGATTGTCTGGGTATTAACCCCTGGCAATATCGGTAAATGAAGAGCGTATTTCGCATG 360
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
DB 361 GTGGTTGACTGTCCGGCGGATAAAACATTCCTCGTATTCGCGTTAAACGGCGGATCG 420
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
DB 421 CTGGAAAAGATCTGCAAGAAAGATATGGCAACCGCGCGGCGGCTGGTGGATCT 480
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
DB 481 GCCATGCGTCATGTCATCATCTCGATCGCTGAACCTTCGATCAGTTCAAAGTCAGCGTG 540
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgIleLeuAlaLysGlnIleAsp 200
DB 541 AAAGCGTCTGACGCTCTCCCTCGCTGTAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValLysSer 220
DB 601 CAGCCGTTGCTATCGGGATCACCGAAGCGCGTGGTGGCGGCGAGCGGGCAGTAAATCC 660
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
DB 661 GCCATTGGATTAGCTCTGCTGTCTGAAGGCATCGGCACACGCTGCGCGTATCGCTG 720
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB 721 GCGGCGCATCGGTGCAAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTGCGTATC 780
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
DB 781 CGTTCGCGAGGATCAACTTCATCGCTGCGCACCTGCTGCGTTCAGGATTTGATGT 840
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
DB 841 ATCGGTACGGTTAACGCGCTGGAGCAACGCTGGAAGATATCATCTCCGATGGACGTT 900
```

```
301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
901 TCGATTATCGGCTGGTGGTGAATGGCCAGGTGAGCGCTGGTTCTTACATCTCGCGCTC 960
321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
961 ACCGGCGGCAACAAGAAAGCGCTCTATGAAGATGGCGTGGCGCAAGACCGTCTGGAC 1020
341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
1021 AACACGATATGATCGACCGTGAAGCACCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080
361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
1081 GAAGCGCTCGAATTCAGCTTCAGCAGGTTGAAGAA 1116

RESULT 2
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US2003009277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragm
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 23, 2003, 18:51:50 ; Search time 402 Seconds
(without alignments)
3024.309 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866

Sequence: 1 MHNQAIQRRKSTRIYGVN.....RAKASQLEARRIDVQVQEK 372

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q=/cg2_1/USPTO.spool/US09921992/runat_21112003_162548_18522/app_query.fasta_1.519
-DB=Published Applications NA -QMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09921992_@CGN_1_1_347_@runat_21112003_162548_18522
-NCPUS=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

- 1: /cg2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cg2_6/prodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cg2_6/prodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cg2_6/prodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cg2_6/prodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cg2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cg2_6/prodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cg2_6/prodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cg2_6/prodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cg2_6/prodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cg2_6/prodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cg2_6/prodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cg2_6/prodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cg2_6/prodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cg2_6/prodata/1/pubpna/US10_NEW_PUB.seq.*
- 16: /cg2_6/prodata/1/pubpna/US60_NEW_PUB.seq.*
- 17: /cg2_6/prodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1866	100.0	1119	9	US-09-921-992-3
					Sequence 3, Appli

2	1603.5	85.9	1830121	14	US-10-329-960-1	Sequence 1, Appli
3	1282	68.7	640681	10	US-09-790-988-1	Sequence 1, Appli
4	822	44.1	1083	10	US-09-974-300-1692	Sequence 1692, Ap
5	808	43.3	1155	14	US-10-156-761-2548	Sequence 2548, Ap
6	808	43.3	9025608	14	US-10-156-761-1	Sequence 1, Appli
7	794	42.6	1180	10	US-09-881-752A-195	Sequence 195, App
8	786.5	42.1	1155	14	US-10-156-761-1636	Sequence 1636, Ap
9	761	40.8	1134	10	US-09-738-626-2205	Sequence 2205, Ap
10	751.5	40.3	3309400	10	US-09-738-626-1	Sequence 1, Appli
11	462	24.8	2520	9	US-09-712-363-110	Sequence 110, App
12	452	24.2	706	14	US-09-921-992-1	Sequence 1, Appli
13	439	24.2	670	9	US-10-066-543-1184	Sequence 1184, Ap
14	339	18.2	670	9	US-09-921-992-22	Sequence 22, Appl
15	326	17.5	596	9	US-09-921-992-23	Sequence 23, Appl
16	321	17.2	584	9	US-09-921-992-21	Sequence 21, Appl
17	292	15.6	33675	9	US-09-921-992-2	Sequence 2, Appli
18	277	14.8	705	9	US-09-921-992-34	Sequence 34, Appl
19	262	14.0	601	9	US-09-921-992-11	Sequence 11, Appl
20	259	13.9	528	9	US-09-921-992-15	Sequence 15, Appl
21	258	13.8	379	9	US-09-921-992-16	Sequence 16, Appl
22	243	13.0	511	9	US-09-921-992-36	Sequence 36, Appl
23	241	12.9	498	9	US-09-921-992-37	Sequence 37, Appl
24	230	12.3	403	9	US-09-921-992-24	Sequence 24, Appl
25	221	11.8	938	9	US-09-921-992-13	Sequence 13, Appl
26	218	11.7	395	9	US-09-921-992-18	Sequence 18, Appl
27	215.5	11.5	464	9	US-09-921-992-33	Sequence 33, Appl
28	214	11.5	432	9	US-09-921-992-14	Sequence 14, Appl
29	212	11.4	395	9	US-09-921-992-17	Sequence 17, Appl
30	210	11.3	564	9	US-09-921-992-35	Sequence 35, Appl
31	198.5	10.6	443	9	US-09-921-992-12	Sequence 12, Appl
32	192.5	10.3	293	9	US-09-921-992-25	Sequence 25, Appl
33	179	9.6	388	9	US-09-921-992-44	Sequence 44, Appl
34	174	9.3	440	9	US-09-921-992-38	Sequence 38, Appl
35	170.5	9.1	211	9	US-09-921-992-45	Sequence 45, Appl
36	166.5	8.9	456	9	US-09-921-992-26	Sequence 26, Appl
37	166	8.8	348	10	US-09-974-300-6102	Sequence 6102, Ap
38	157.5	8.4	869	9	US-09-921-992-42	Sequence 42, Appl
39	155.5	8.3	551	9	US-09-921-992-41	Sequence 41, Appl
40	155	8.3	430	9	US-09-921-992-29	Sequence 29, Appl
41	155	8.3	615	9	US-09-921-992-6	Sequence 6, Appli
42	154.5	8.3	619	9	US-09-921-992-27	Sequence 27, Appl
43	151	8.1	422	9	US-09-921-992-28	Sequence 28, Appl
44	150	8.0	589	9	US-09-921-992-7	Sequence 7, Appli
45	149.5	8.0	399	9	US-09-921-992-47	Sequence 47, Appl

RESULT 1

US-09-921-992-3
; Sequence 3, Application US/09921992
; Patent No. US2002008946A1
; GENERAL INFORMATION:
; APPLICANT: Boronat, Albert;
; APPLICANT: Campos, Narciso;
; APPLICANT: Rodriguez-Concepcion, Manuel;
; APPLICANT: Rohmer, Michel;
; APPLICANT: Seeman, Myriam;
; APPLICANT: Valentin, Henry E.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; APPLICANT: Venkatesh, Mylavara
; TITLE OF INVENTION: Methvl-D-Erythritol Phosphate Pathway Genes
; FILE REFERENCE: 1616.107/35-21(51897)US
; CURRENT APPLICATION NUMBER: US/09/921,992
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/223,483
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 3
; LENGTH: 1119
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:

QY 226 LeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProVal 245
Db 694 GGGTTATTGATGAAGGCAITGGCGATACGATCGTATTTCGCTTGTCTGTAACCTGAA 753
QY 246 GluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIle 265
Db 754 GATGAATCAAGATCGGTTTGTATCTTAAATCGCTTGGCCTACGCTTCTAACGGGTATC 813
QY 266 AspPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsn 285
Db 814 AACTTATTGCTTGTCCAAAGTGTCTCGCCAGAAATTTAAACGTGATTCAGGTGATCGAG 873
QY 286 AlaLeuGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCys 305
Db 874 GCTTTAGAGAGCGTTTAGAAGATATCCGTACCAATGACCGTTTCGGTTATTGGCTGT 933
QY 306 ValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLys 325
Db 934 AAGGTAATGGCCAGGTGAAGCAAAAGAACGAGATATCGGGTGTGTGGGGCTGCGCCT 993
QY 326 LysSerGlyLeuTyrgluAspGlyValArgLysAspArgLeuAspAsnAspMetIle 345
Db 994 CGTTCATGCTTATCGTAATGCTGAGAAAGCCATTTAATTGATACCAATCAATTTGGIT 1053
QY 346 AspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIle 365
Db 1054 GATGAATCGAACTATGTTGCTCAACGTTTCAAGAGCTTGAAGAAGCTTAATCTAAA 1113
QY 366 AspVal 367
Db 1114 GAAATT 1119

RESULT 15

US-08-827-190-9/c

Sequence 9, Application US/08827190
Patent No. 5859367
GENERAL INFORMATION:
APPLICANT: Rather, Philip N.
TITLE OF INVENTION: Methods For Screening For Antimicrobials
TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/827,190
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 886 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-827-190-9

Alignment Scores:
Pred. No.: 2,68e-84 Length: 886
Score: 812.00 Matches: 158
Percent Similarity: 70.65% Conservative: 49
Best Local Similarity: 53.92% Mismatches: 86
Query Match: 43.52% Indels: 0
DB: 2 Gaps: 0

US-09-921-992-50 (1-372) x US-08-827-190-9 (1-886)

QY 30 ValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIleLys 49
Db 881 ATCCAAGCATGACAAACAAACACATGATGTAGAAGACCGTTCGCGAAATTAAC 822
QY 50 AlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAlaAla 69
Db 821 CGTTTCGGCTGAAGCCCGGATGCCAAATCGTTCCGGTAGCATGTCCGGATGAACCGCGGCA 762
QY 70 GluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAspIleHisPhe 89
Db 761 AACGCCATTCCGGATATTAAAGCCGATTCCTTCCTCTGTTGTCATACATTC 702
QY 90 AspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnPro 109
Db 701 GATTATAAACTTCGCTTGAAGCCATTGAAGCGCGGCAGATAAAATCCGAATCAACCCC 642
QY 110 GlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLysAsn 129
Db 641 GGCAATATCGCGCGCGCGGCGGAAAGTTGAAGCGGTTGTTAAAGCGGCCAAAGCAAGGC 582
QY 130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnLysTyr 149
Db 581 ATTCCGATCAGAAATCGAGTAAACCGCGGTTTCATTGGAAGAAACGGATTTTAGAATAATAC 522
QY 150 GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAsp 169
Db 521 GGTATCCGACTGCCGATGGAATGTTAGAAAGCGCACTTCATCACATTAATAATTTTGTAG 462
QY 170 ArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaVal 189
Db 461 GATCTTGATTTTTCAGCATATTATGTCAGCATGAAGCCCTCTGACGTGAACCTTGAATC 402
QY 190 GluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThrGlu 209
Db 401 GAGGCTTATGAAGAACGACGAAAGCGTTTGACTACCCGCTTCACCTCGGATCACCGAG 342
QY 210 AlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuSer 229
Db 341 TCAGGAACACTGTTTCCGCGCACAGTAAAGCGCAGCAGGACTCGCGCCCATTTTAAGC 282
QY 230 GluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIleLys 249
Db 281 AAAGGCATCGGGAACACCATTCGCGCATTTTCACTAAGCGCAGACCCCTGTAGAGAGTAAA 222
QY 250 ValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIleAla 269
Db 221 GTAGCAAGGAGCTTCTGAATCTTTTCGGCTTAGCTTCAATGTGCGCCACGCTCTCTCA 162
QY 270 CysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGluGln 289
Db 161 TGCCCGACTTTCGCGCGGATTGAGATTGATCTAATCAGCATTCGCAATGAAGTGAAGAG 102
QY 290 ArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsnGly 309
Db 101 TATATTTCTAAGATAAAAGCGCGGATTAAAGTTGCTTCTCGGCTGCGCTGTAACCGCA 42
QY 310 ProGlyGluAlaLeuValSerThrLeuGlyValThrGly 322
Db 41 CCTGTGAAGCGGAGAGAGCTGATATCGAATTCGCGCGC 3

Search completed: November 23, 2003, 19:44:10
Job time : 952 secs

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-170-187-10

Alignment Scores:
Pred. No.: 111e-146 Length: 886
Score: 1353.00 Matches: 268
Percent Similarity: 96.26% Conservative: 15
Best Local Similarity: 91.16% Mismatches: 11
Query Match: 72.51% Indels: 0
DB: 4 Gaps: 0

US-09-921-992-50 (1-372) x US-09-170-187-10 (1-886)

QY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
DB 884 GCTGTCCTATCTATGACGATATCGCCACGACGAGATGTTGAGCCCACTGTGGCGCAATTC 825
QY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
DB 824 CAATCACTTGAGCGTGTAGGTGTGATATCGTCCGCGTGTCTGTTCTCTACGATGATGCA 765
QY 69 AlaGluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAspIleHis 88
DB 764 GCAGAAAGCCTTTAAATTAATTAAGCAGCGCGTGAATGTGCCATTTGGTTGCGGATATTCAC 705
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
DB 704 TTTGACTACCGTATCCGATGAAAGTGGCTGAATATGTTGTTGACTGCTCAGCAATTAAC 645
QY 109 ProGlyAsnIleGlyAsnGluArgIleArgMetValValAspCysAlaArgAspLys 128
DB 644 CCAGGTAATATCGCAGTGAAGAGCGTATTCCGCAAGTGTGTCATAGTGTCTGTCATCAC 585
QY 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGluLys 148
DB 584 AACATTCCTATCCGTATAGGGGTCAATGGCGGGTCACTGGGAAAAGATATCCAGAAAGAAA 525
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 524 TACGGTGAGCCAAACACTGAGGACATTTGGTTGAATCAGCAATGCGACATGTTGATATCTTG 465
QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
DB 464 GACAGCGTGAATTTCCGATCAGTTCAGGTCAAGTCAAGTCAAGCGTGTCTTCTTCTGCC 405
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
DB 404 GTCGGCTCTTATCGTTTATTGGCGCAAAAATTTGATCAACCACTTCACCTCGGTATTACA 345
QY 209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
DB 344 GAAGCGGTGGGCTCGTCTGTTTCAAGTGAATCAGCAATTCGTTGTTGATGTTGTTG 285
QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
DB 284 GCTGAAGGTATCGGCATACGTTACGTATCTACTCGCGCAGATCTCTGTTGAGGAAGTG 225
QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
DB 224 AAAGTCGGTTTGTATATTCCTAAATTCGTTACGATCGCTCAGCTCGGCATCAACTTTATT 165
QY 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
DB 164 GCTTGCCCAACCTGTTTCCGCCCAAGAAATTTGATGTGATGTTGATGTTGTTGAG 105
QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleLeuGlyCysValValAsn 308
DB 104 CAGCGCTCGAAGATATATCAGCGCGATGATGCTCTATTATTGTTGTTGTTGATGAAT 45
QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGly 322
DB 44 GGCCCGGGTGAAGCCGAGGTTTCTACTTTTAGGTGTGGCTGCG 3

RESULT 14

US-09-328-352-3780
Sequence 3780, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 3780
LENGTH: 1137
TYPE: DNA
ORGANISM: Acinetobacter baumannii
US-09-328-352-3780

Alignment Scores:
Pred. No.: 2.67e-128 Length: 1137
Score: 1195.00 Matches: 231
Percent Similarity: 81.77% Conservative: 65
Best Local Similarity: 63.81% Mismatches: 66
Query Match: 64.04% Indels: 0
DB: 4 Gaps: 0

US-09-921-992-50 (1-372) x US-09-328-352-3780 (1-1137)

QY 6 ProfileGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGly 25
DB 34 CCAATTAACCGTCGACCAACACGATAAATCCGTGTTGGTTCGTCTATGTCGGTGGCGAT 93
QY 26 AlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAlaThrVal 45
DB 94 GCACCTATTAGTGTGCAAGATGACAAATACCGAACTTGGCATGTTGCCGCACTGTG 153
QY 46 AsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThr 65
DB 154 GCTCAGATTGACGGTGTGTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 213
QY 66 MetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsnValProLeuValAla 85
DB 214 ATGGAAGCTGCTGAAGCCCTTTGGTCAATCCGATCCGATGTTTTCAGTTCCATTAGTAGCT 273
QY 86 AspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeu 105
DB 274 GATATCCATTTTGACCATAGAAATGCTTTGGCAGTTGACGATTTATGTCGGACTGCTTG 333
QY 106 ArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAla 125
DB 334 CGTATTAAACCCGGCAATATCGTTTCAGACCAAGAAAGTTTCGTGAAGTTTCGTGGCTGGCA 393
QY 126 ArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeu 145
DB 394 CGTCATCAGCGTATTTTATGCGTATGTCGCGTGAATTCGCGGTCTCTAGAAAAAGATTTA 453
QY 146 GlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisVal 165
DB 454 CAGAAAAATATGGCAGCTACAGGGCAGGCACTTCTTTCAGTCAGCTTACGTCATATT 513
QY 166 AspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspVal 185
DB 514 GATATTTAGACCGTCTTGACTTCCATGAGTTTAAAGTCAGTGAAGATCAAAATGTG 573
QY 186 PheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeu 205
DB 574 TTTTAAACCATGGATGCTTATGCTTCTCTCAACAAATGATATCCATTACACCTT 633
QY 206 GlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGly 225
DB 634 GGAGTGACTGAAGCTGGTATTTTACCGTACAGGTACTGTGAAATATCCGATTTGCTTTGTT 693

APPLICANT: Rather, Philip N.
TITLE OF INVENTION: Methods For Screening For Antimicrobials
UTILIZING aarC And Compositions Thereof
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/827,190
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 886 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

US-08-827-190-10

Alignment Scores:

Pred. No.: 1,11e-146 Length: 886
Score: 1353.00 Matches: 268
Percent Similarity: 96.26% Conservative: 15
Best Local Similarity: 91.16% Mismatches: 11
Query Match: 72.51% Indels: 0
DB: 2 Gaps: 0

US-09-921-992-50 (1-372) x US-08-827-190-10 (1-886)

QY 29 AlaValGlnSerMetThrAsnThrArgThrValAspValGluAlaThrValAsnGlnIle 48
DB 884 GCTGTCCTAATCTATGACGAATACGCGCAGCGGATGTTGAAGCCACTGTGCGCAATC 825
QY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
DB 824 CAATCACTTGAAGCGTGTAGGTGTTGATATGTCGCCGCTGTCTGTTCTACGATGATGCA 765
QY 69 AlaGluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAspIleHis 88
DB 764 GCAGAGCCCTTAATTAATTAAGCAGCGCGGATGTCCTGCTGCTGCGGATATCATC 705
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
DB 704 TTTGACTACCGTATCGCATGAAGGCTGAATATGTTGCTGACTGCTACGAATTAAC 645
QY 109 ProGlyAsnIleGlyAsnGluArgIleArgMetValValAspCysAlaArgAspLys 128
DB 644 CCAGGTAAATATCGGCAAGTGAAGGCGTATTCGCCAAGTGTGATAGTCTGCTCATCAC 585
QY 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnLys 148
DB 584 ACATTCCTATCCGTATAGGGTCAATGCGGGTCACTGGAAGAAGATATCCAGAAA 525
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 524 TACGGTGACCAACACCTCAAGCATTTGTTGAATCAGCAATCGCATGTTGATATCTTG 465

QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
DB 464 GACAGCTGAATTTGATCATGTTCAAGTCAGTGTAAAGCGTCGGATGCTCTTCTTGCC 405
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
DB 404 GTCCGCTCTTATCGTTTATTGCGCAAAAAATTGATCAACCACTTCACCTCGGTATTACA 345
QY 209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
DB 344 GAAGCGGGTGGGGCTCGTTCTGTTTCAAGTGAATCAGCAATTTGGTCTTGGTATTGTTG 285
QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
DB 284 GCTGAAGGTATCGCGATACGTATCTCCTCCTCGCGCAGATCTCTGTTGAGGAAGTG 225
QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
DB 224 AAAGTCGGTTTTGATATTTCTAAATCGTTACGGATCCGCTCAGTGGCATCACTTTATT 165
QY 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
DB 164 GCTTCCCAACCTGTTTCAGCCCAAGAAATTTGATGTGATGTTACGTAATGCTTTGGAG 105
QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
DB 104 CAGCGCTCGAAGATATTATCAGCGCGATGATGTTCTATTATTGTTGTTGTAGTGAAT 45
QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGly 322
DB 44 GCGCCGGGTGAAGCGGAGGTTTCTACTTTAGGTGGCTGGC 3

RESULT 13

US-09-170-187-10/c
Sequence 10, Application US/09170187
Patent No. 6383745
GENERAL INFORMATION:
APPLICANT: Rather, Philip N.
TITLE OF INVENTION: Methods For Screening For Antimicrobials
UTILIZING aarC And Compositions Thereof
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/170,187
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/827,190
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 886 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

QY 201 GlnProLeuHisLeuGlyLeuThrGluAlaGlyValAlaArgSerGlyAlaValLysSer 220
DB 1527 CAGCCCTGCACCTGGGCATCACCGAGGCGCGGCTCGCTCCGGCACGGTGAAGTCG 1468
QY 221 AlaileGlyLeuGlyLeuLeuSerGluGlyLeuGlyAspThrLeuArgValSerLeu 240
DB 1467 GCGGTGGGCTGGGCATGCTCTGGCGGAGGAATCGGCACACCATCCGGATTTCCCTG 1408
QY 241 AlaAlaAspProValGluGluLeuLeuValGlyPheAspLeuLeuSerLeuArgile 260
DB 1407 GCTGCCGATCCCGTCAGAGATCAAGTTCGGTTTCGACATCTCAAGTCCCTGACCTG 1348
QY 261 ArgSerArgGlyLeuLeuPheAlaCysProThrCysSerArgGlnGluPheAspVal 280
DB 1347 CGCTCCCGTGGCATCAACTTCATCGCTGCCCGAGCTGTTCCGGCGAGAACTTCGACGTG 1288
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleLeuThrProMetAspVal 300
DB 1287 GTCAAGACCATCAACGAGCTGGAAGGCGCTGGAGGACCTGCTGTGTCGATGGACGTG 1228
QY 301 SerIleileGlyCysValValAlaGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
DB 1227 GCGGTGATCGGTTCGTCAACGTCGCGGCGGAAGGAGGAGGAGGAGGAGGAGGAGGAG 1168
QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrgluAspGlyValArgLysAspArgLeuAsp 340
DB 1167 ACCGGCGGCACCTCCGAACCTG---GTGTATATCGACGGCAAGCGCTCGCAGAACTGACC 1111
QY 341 AsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
DB 1110 AACGAAACCTGTGTGACGAGCTGGAACGCGCTGATCCGCGAGGAGGCGCGAGAGGCGC 1051
QY 361 GluAla 362
DB 1050 GAGGCC 1045

RESULT 11

US-09-252-991A-6339
; Sequence 6339, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6339
; LENGTH: 2088
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6339

Alignment Scores:

Pred. No.:	2,47e-147	Length:	2088
Score:	1363.50	Matches:	266
Percent Similarity:	86.59%	Conservative:	44
Best Local Similarity:	74.30%	Mismatches:	47
Query Match:	73.07%	Indels:	1
DB:	4	Gaps:	1

US-09-921-992-50 (1-372) x US-09-252-991A-6339 (1-2088)

QY 5 AlaProIleGlnArgArgLysSerThrArgIleTyrgluValGluAsnValProIleGlyAsp 24
DB 5 TCTCCGATCATTCGCCCGCAAGTCTCGGAAATCTGGGTTCGCAACGTCCTCGGTGGCGGC 64

QY 25 GlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThr 44
DB 65 GACGCCCGATCGCGTTCGAGCATGACCAACACCGAGACCTGCGACCTGCTGCCACC 124
QY 45 ValAsnGlnLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValPro 64
DB 125 GTCCGCGAGATCCCGCGCTGGAAGATCCCGCGCGGACATCGTCCGGGTTTCCGTCCC 184
QY 65 ThrMetAspAlaAlaGluAlaPheLysLeuLysGlnGlnValAsnValProLeuVal 84
DB 185 GACATGGACCGCGCGAGCATTCGGCAAGATCAAGCAGCAGGTCAACGTCGCGTGGTC 244
QY 85 AlaAspIleHisPheAspTyrgluAlaLeuLysValAlaGluTyrglyValAspCys 104
DB 245 GCGGACATCCATTCGACATTCGATTCGCTCGCGCTGCGGAGCTGGGAGTGACATGC 304
QY 105 LeuArgileAsnProGlyAsnIleGlyAsnGluArgileArgMetValValAspCys 124
DB 305 CTGCGCATCAATCCGGGCAACATCGGTCCGAGGACCGGGTCAAGCGCGTGTTCATGCC 364
QY 125 AlaArgAspLysAsnIleProIleArgileGlyValAsnAlaGlySerLeuGluLysAsp 144
DB 365 GCGCGCGAGCGCAACATCCGATCGTATCGCGCTCAATGCGGTTCGCTGGAAGGAC 424
QY 145 LeuGlnGluLysTyrglyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHis 164
DB 425 CTGCAGAGAAATACGGGAAACCGGACCGCGGAGCCCTGCTCGAATCGGCCATCGCCAC 484
QY 165 ValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAsp 184
DB 485 GTCGATCATCTCGCAAGCTGGACTTCCAGAACTTCAAGGTCAAGGTCAAGGCTCCGAC 544
QY 185 ValPheLeuAlaValGluSerTyrgluLeuAlaLysGlnIleAspGlnProLeuHis 204
DB 545 GTCTTCATGGCGCTCGCGCTATCGCTGTGTCGCGCAGCAGATCGAGCAGCCCTGCAC 604
QY 205 LeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224
DB 605 CTGGGATCACCGAGCGCGCGCTCGGCTCGGCAACGTCGCGTGGGCGTGGGCTG 664
QY 225 GlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspPro 244
DB 665 GGCATGCTCTCGCGGAGGAAATCGCGACATCCCGATTTCCCTGGCTGCCGATCCG 724
QY 245 ValGluGluLysValGlyPheAspIleLeuLysSerLeuArgileArgSerArgGly 264
DB 725 GTCCGAGAGATCAAGTTCGATTCGACATCTCAAGTCCCTGCACCTCGCTCCCGTGGC 784
QY 265 IleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrVal 284
DB 785 ATCAACTTCATCGCTCGCGAGCTGTTTCGCGGCAAGCTTCGACGTGTGTAAGACCATG 844
QY 285 AsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGly 304
DB 845 AACGAGCTGGAAGGCGCGCTGGAGGACCTGCTGTGTCGATGGACGTGCGCGTGCATCGGT 904
QY 305 CysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsn 324
DB 905 TCGCTGCTCAACGCTCGCGGCAAGCAAGGAGGCGCCATGTCGGCTCCCGCGGCACT 964
QY 325 LysLysSerGlyLeuTyrgluAspGlyValArgLysAspArgLeuAsnAsnAspMet 344
DB 965 CCGAACCTG---GTGTATATCGACGCAAGCGCTGCGAGAACTCCCAACGCAACCTG 1021
QY 345 IleAspGlnLeuGluAlaArgileArgAlaLysAlaSerGlnLeuAspGluAla 362
DB 1022 GTGGACGAGCTGGAACGCTGATTCGCGCAAGGCGCGGAGGAGGAGGAGGAGGAGG 1075

RESULT 12

US-08-827-190-10/c
; Sequence 10, Application US/08827190
; Patent No. 5858367
; GENERAL INFORMATION:

```
Db 535 ATTATAGCGCTTCTCCGATCATTCGCGCAAGTCTCGGAAATCTGGGTGCGCAACGTC 594
Qy 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 595 CCGGTGGGGCGGCGACCGCGCGATCGCGGTGCGAGAGCATGCCAACACCGGAGCTCGAC 654
Qy 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
Db 655 GTGCTGCCACCGTCCGCGAGATCCCGCGCTGGAAGATCGCGCGCGGACATCTGTGCGG 714
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsn 80
Db 715 GTTTCGCTCCCGACATGACGCGCGCGGAGGATTCGCAAGATCAAGCAGCAGGTCAAC 774
Qy 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db 775 GTGCCCGCTGGTCCGACATCCACTTCGACTATCGCATCGCCCTGCGCGTCCGCGAGCTG 834
Qy 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgMet 120
Db 835 GGAGTGGATGCTGCGCATCAATCCGGGCAACATCGTCCGAGGACCGGGTCAAGGCC 894
Qy 121 ValValAspCysAlaAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
Db 895 GTGGTTCGATCGCGCGCGGCGCAACATCCCGATCCGATCGGTATCGCGTCAATGCCGGTTCG 954
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGluSer 160
Db 955 CTGGAAAGAGGACTGCGAGAGAAATACGCGCAACCGACCCCGGAGCCCTGTCTGAATCG 1014
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
Db 1015 GCCATGCGCACGTCGATCATCTCGCAAGCTGGACTTCAGAACTTCAAGGTCAAGCGTC 1074
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 1075 AAGGCTCCGACGCTTTCATGCGCGTCCGCGCTATCGCTGCGCGAGGAGATCGAG 1134
Qy 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220
Db 1135 CAGCCCTTCACCTGGGACATCACGAGGCGCGCGCTCGCTCGCGCACGCTGAAGTCG 1194
Qy 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
Db 1195 GCGGTGGGCTGGGATGCTCTGCGCGAGGAATCGCGACACCATCCGGAATTTCCCTG 1254
Qy 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 1255 GCTGCCGATCCGTCGAGGAGATCAAGGTGCGTTCGACATCCTCAAGTCCCTGCACCTG 1314
Qy 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnPheAspVal 280
Db 1315 CGCTCCGCTGGGATCAACTTCACTGCTGCGCGAGGCTGTTCGCGGAGAACTTCGACGTG 1374
Qy 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
Db 1375 GTGAAGACCATGACGACTGGAAGGCGCCCTGGAGGACCTGCTGTGTCGATGACGCTG 1434
Qy 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
Db 1435 GCCGTGATCGGTTCGTCGTCGTCGCGGAGGAGCAAGGAGGCCCATCTCGGCGCTC 1494
Qy 321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
Db 1495 ACCGGCGGCACTCCCAACCTG---GTGTATATCGACGCGCAAGCCGTCGAGAACTGACC 1551
Qy 341 AsnAsnAspMetIleAsnGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
Db 1552 AACGACACCTGCTGGACGAGCTGGAACCGGTGATCCCGCTGATCCCGCAGAGGCGCGAGAGGCC 1611
Qy 361 GluAla 362
|||||
```

```
Db 1612 GAGGCC 1617
RESULT 10
US-09-252-991A-6131/c
; Sequence 6131, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6131
; LENGTH: 2202
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6131
Alignment Scores:
Pred. No.: 2,46e-148 Length: 2202
Score: 1372.50 Matches: 267
Percent Similarity: 86.46% Conservative: 46
Best Local Similarity: 73.76% Mismatches: 48
Query Match: 73.55% Indels: 1
DB: 4 Gaps: 1
US-09-921-992-50 (1-372) x US-09-252-991A-6131 (1-2202)
Qy 1 MethAsnGlnAlaProIleGlnArgLysSerThrArgIleTyrValGlyAsnVal 20
Db 2127 ATTATAGCGCTTCTCCGATCATTCGCGCAAGTCTCGGAAATCTGGGTGCGCAACGTC 2068
Qy 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 2067 CCGGTGGGCGGCGAGCGCGCGATCGCGTGCAGAGCATGCCAACACCGAGACCTGCGAC 2008
Qy 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
Db 2007 GTGCTGCGCACCGTCCGCGAGATCCGCGCTCGGAAGATCGCGCGCGGAGATCGTGGG 1948
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsn 80
Db 1947 GTTTCGCTCCCGACATCGACGCGCGCGGAGCATTCGCAAGATCAACGAGCAGGTCAAC 1888
Qy 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db 1887 GTGCGGTGGTGGCGGACATCCACTTCGACTATCGCATCGCCCTGCGCGTCCGCGAGCTG 1828
Qy 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
Db 1827 GGAGTGGACTGCTCGCATCAATCCGGGCAACATCGGTTCGCGAGGACCGGGTCAAGGCC 1768
Qy 121 ValValAspCysAlaAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
Db 1767 GTGTCGATGCGCGCGCGGAGCGCAACATCCCGATCCGATCCGTCGCGGTCAATGCCGTTG 1708
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
Db 1707 CTGGAAAGAGGACTGCGAGAGAAATACGCGCAACCGACCCCGGAGACCTGCTGCAATCG 1648
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
Db 1647 GCCATGCGCACGCTCGATCATCTCGCAAGCTGGACTTCCAGAACTTCAAGGTCAAGCGTC 1588
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 1587 AAGGCTCCGACGCTTTCATGCGCGCTCGCCCGCTATCGCCCTGCTGGCAGGACGATCGAG 1528
```

STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/170,187
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/827,190
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1079 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-09-170-187-7

Alignment Scores:
Pred. No.: 2,76e-174 Length: 1079
Score: 1593.50 Matches: 317
Percent Similarity: 94.93% Conservative: 20
Best Local Similarity: 89.30% Mismatches: 17
Query Match: 85.40% Indels: 1
DB: 4 Gaps: 1

US-09-921-992-50 (1-372) x US-09-170-187-7 (1-1079)

QY 7 IleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAla 26
DB 6 ATTAAGCGTGTGAATCGACAAATTTATGTGGAAATGTACCAATTTGGTGGGATGCG 65
QY 27 ProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsn 46
DB 66 CCTATTGCCGTGCAATCAATGACAAATPACTCGCACCACTGATGTGGAAGCGACAGTTGCT 125
QY 47 GlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMet 66
DB 126 CAAATTAAATCATTTAGAACGTGTGGTGACAGATTTGTCGTATCTGTTCACCAATG 185
QY 67 AspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaAsp 86
DB 186 GATGCTCGGAGCATTTAAACAAATTAACACAACTGAATGTTCCGCTCGTAGCAGAT 245
QY 87 IleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArg 106
DB 246 ATTCATTTCGACTATCGTATCGGCTTAAAGTCGACAGATATGAGTGGATTTGTTACGT 305
QY 107 IleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArg 126
DB 306 ATCAATCCTGGCAACATTTGGTGTGAAGATCGCGTCCGCTGTTGTATGTGGCGCA 365
QY 127 AspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGln 146
DB 366 GACAAAATATTCGATTCGATTTGGTAAATGCAGGCTCTTTAGAAAAGATTTGCAA 425
QY 147 GluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAsp 166
DB 426 GAAAAATATGCGGAACCAACGCGCAAGACCTTGTAGAAATCCGCAATTCGCTCATGTAGAA 485

QY 167 HisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPhe 186
DB 486 ATTTAGATCGTCTTAACTTCGATCAGTTTAAAGTGAGCGTAAAGCCTCCGATGTATTC 545
QY 187 LeuAlaValGluSerTyrArgIleLeuAlaLysGlnIleAspGlnProLeuHisLeuGly 206
DB 546 TTAGCGGTTGAATCTATCTGTTTACTGGCTAAGCAATTAACACGCTTTACATTTAGCG 605
QY 207 IleThrGluAlaGlyGlyValAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
DB 606 ATTACAGAACAGGTGGCGCACCGCTGGTGAGTAAATCTGCAGTGGGTTTAGGAATG 665
QY 227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
DB 666 TTATTAGCTGAGGCGCATTTGGCGATACACTACGCGTCTCTTTGGCGGACAGATCCTGTAGAG 725
QY 247 GluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsn 266
DB 726 GAAATCAAAGTCGGTTTGTATTTTGAATCTTTACGGATTCGTTCAAGAGAAATTAAC 785
QY 267 PheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAla 286
DB 786 TTTATTTGCTTGCCCAACCTGTTCTGCCAAGAAATTTGATGTAATCGGTACAGTAATGCG 845
QY 287 LeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysVal 306
DB 846 CTAGAACCAACGCTTGAAGATATTATTACCAATGGATGTATCTATTATCGTTGTGTA 905
QY 307 ValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyAsnLysLys 326
DB 906 GTGAATGGTCTTGGCGAGGACCTCGTCTCCGATCTCGGCGTAACGGCGGTAAACAAAAA 965
QY 327 SerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAspMetIle 345
DB 966 ACGGTTATTTATCTTGACGAGAACCCCAAAAGAGCGTTTGTATACGAGATATAGTG 1025
QY 346 AspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
DB 1026 AACCAATTAGAAAGCAAAATTCGTGCGAAAGTCGACGACCAAGAT 1070

RESULT 9

US-09-252-991A-6420
; Sequence 6420, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6420
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6420

Alignment Scores:
Pred. No.: 1,58e-148 Length: 1644
Score: 1372.50 Matches: 267
Percent Similarity: 86.46% Conservative: 46
Best Local Similarity: 73.76% Mismatches: 48
Query Match: 73.55% Indels: 1
DB: 4 Gaps: 1

US-09-921-992-50 (1-372) x US-09-252-991A-6420 (1-1644)

QY 1 MetHisAsnGlnAlaProIleGlnArgLysSerThrArgIleTyrValGlyAsnVal 20

Db 391812 GTAAATGCGCTAGAACACCGCTTGAAGATATTATTACCAATGATGATCTATTATC 391871
QY 304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323
Db 391872 GGTGTGTGTGTGAATGCTGCGGAGGACCTGCTCCGATCTCGCGTAACGGCGGT 391931
QY 324 AsnLysLysSerGlyLeuValGluAspGlyValArg---LysAspArgLeuAspAsnAsn 342
Db 391932 AACAAAAAGCGGTATTATCTTGACGGAGACGCCAAGAGCGTTTGTATACGAA 391991
QY 343 AspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
Db 391992 GATATAGTGAACCAATTAGAGCAAAAATTCGTGCGAAAGTCCGACGACAAAGTCCAAA 392051
QY 363 ArgArgIle 365
Db 392052 AACAGAAATT 392060

RESULT 6

US-09-643-990A-1

Sequence 1, Application US/09643990A

Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann

Mark D. Adams

Owen White

Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Alignment Scores:

Pred. No.: 1.38e-170

Score: 1603.50

Length:

Matches:

1830121

320

Percent Similarity: 93.66% Conservative: 20
Best Local Similarity: 88.15% Mismatches: 22
Query Match: 85.93% Indels: 1
DB: 4 Gaps: 1
US-09-921-992-50 (1-372) x US-09-643-990A-1 (1-1830121)
QY 4 GlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23
Db 390972 CAGCCAACCTATTAAAGCTCGTAATCGACAAAATTTATGTGGAAATGTACCAATTGTT 391031
QY 24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAla 43
Db 391032 GGGGATGCGCTATTGCGGTGCAATCAATGACAAATACTCGCACCACCTCATGTGAAGCG 391091
QY 44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63
Db 391092 ACAGTTGCTCAAAATTAATCATTAAGACGTGTGTGGCAGATATTGTCGTATCTGTT 391151
QY 64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeu 83
Db 391152 CCACCAATGGATGCTGCGGAGCATTTAAACAAATTAACACAACTGAATGTTCCGCTC 391211
QY 84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103
Db 391212 GTAGCAGATATTCAATTCGACTATCGTATCGCGTTAAAAAGTCGACAAATATGAGTGGAT 391271
QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123
Db 391272 TGTTCAGCTATCAATCTCGGCAACATTTGCTGTAAGATCGCGTCGCTGCCGTGTTGAT 391331
QY 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143
Db 391332 TGTGCGGAGACAAAAATATTCCGATTCTGATTGTTGTTAAATGCAGGCTCTTTAGAAAA 391391
QY 144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163
Db 391392 GATTTGCAAGAAAAATATGCGAACCCACGCGCAGAACGCTTGTAGAAATCCGATTCGCT 391451
QY 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183
Db 391452 CATGTAGAAAATCTTAGATCTCTTAACCTCGATCAGTTTAAAGTAGCGGTAAAGGCTCC 391511
QY 184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203
Db 391512 GATGATTTCTTAGCGGTGAAATCTTATCGTTTACTGGCTAAACCAATTAACAGCCTTTA 391571
QY 204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223
Db 391572 CATTAGGCATTACAGAACGAGTGGCGCACGCGCTGGTGCAGTAAAAATCTGCAGTGGGT 391631
QY 224 LeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243
Db 391632 TTAGGAATGTTATTAGCTAGGCGCATTTGCGGATACACTACGCGTCTCTTTGGCGGAGAT 391691
QY 244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263
Db 391692 CCTGTAGAGAAATCAAGTCGTTTGTATATTGAAATCTTTACGGATTCGTTCAAGA 391751
QY 264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGlnPheAspValIleGlyThr 283
Db 391752 GGAATTAACCTTTATGCTTGGCCCAACCTGTTCTCGCCCAAGAAATTCATGTAATCGGTACA 391811
QY 284 ValAsnAlaLeuGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303
Db 391812 GTAAATGCGCTAGAACCAACGCTTGAAGATATTATTACCAATGATGATCTATTATC 391871
QY 304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323
Db 391872 GGTGTGTGTGTGAATGCTGCGGAGGACCTGCTCTCGATCTCGCGTAACGGCGGT 391931
QY 324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsnAsn 342
Db 391932 GGTGTGTGTGTGAATGCTGCGGAGGACCTGCTCTCGATCTCGCGTAACGGCGGT 391991

```
Db 775 CAACCACTTCACCTCGGTATTACAGAACGGGTGGGCTCGTTCTGTTCAAGTCAATCA 834
Qy 221 AlaileGlyLeuGlyLeuLeuSerGluGlyLeuGlyAspThrLeuArgValSerLeu 240
Db 835 GCAATGGTCTTGTTGGTGAAGGATCGCGCATACGTTACGTATCTCACTC 894
Qy 241 AlaAlaAspProValGluGlyLeuLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 895 GCGGAGATCCTGTTGAGGAAGTGAAGTCGGTTTGGATTTCTAAATCGTTACGGATC 954
Qy 261 ArgSerArgGlyIleAsnPhelIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
Db 955 CGCTACGTGCGCATCACTTATTGCTTGCCCAACCTGTTTCACGCCCAAGAAATTTGATGTG 1014
Qy 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
Db 1015 ATTGGTACGTAATGCTTTGGAGACGCGCTCGAAGATATTATCACGCGATGGATGC 1074
Qy 301 SerIleileGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
Db 1075 TCTATTATTGTTGTGTAGTGAATGCCCGGTGAAGCCGAGGTTTCTACTTTAGGTGTG 1134
Qy 321 ThrGlyClyAsnLysLysSerGlyLeuTyrgluAspGlyValArg---LysAspArgLeu 339
Db 1135 GCTGGCGCCGAAACCAAAAGTGGTTCTATGAAGATGGCGTTTCGCAAAAAGAGCGTTT 1194
Qy 340 AspAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeu 359
Db 1195 GATAATGACAAATATTATGATGACGTGAGCGGAAATTCGCGCAAAAGCAGCAATGCTT 1254
Qy 360 AspGluAlaArgIleAspValGlnGlnValGlu 371
Db 1255 GATGAAT-AACCGTATAAGATAAACCAAGTGC 1289

RESULT 5
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
;
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/476,102
; FILING DATE: JUN-5-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB186P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 301-309-8504
; TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1830121 base pairs
```

```
;
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.: 1,38e-170 Length: 1830121
Score: 1603.50 Matches: 320
Percent Similarity: 93.66% Conservatives: 20
Best Local Similarity: 88.15% Mismatches: 22
Query Match: 85.93% Indels: 1
DB: 4 Gaps: 1

US-09-921-992-50 (1-372) x US-09-557-884-1 (1-1830121)
Qy 4 GlnAlaProIleGlnArgGlySerThrArgIleTyrglyValGlyAsnValProIleGly 23
Db 390972 CAGCCAACTATTAAAGCGTCTGAATCGACAAAATTTATGTGGAAATGTACCAATGGT 391031
Qy 24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAla 43
Db 391032 GGGGATGCGCCCTATTTCCTGCAATCAATGACAAATACTCGCACCACTGATGTGAAGCG 391091
Qy 44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63
Db 391092 ACAGTTGCTCAATTAATCATTTAGAACGTTGTGGTCAGATATTGTTCTGTATCTGTT 391151
Qy 64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeu 83
Db 391152 CCAACAATGGATGCTGCGGAAGCATTTAAACAAATTTAAACAACAAGTGAATGTTCCGCTC 391211
Qy 84 ValAlaAspIleHisPheAspTyrglyIleAlaLeuLysValAlaGluTyrglyValAsp 103
Db 391212 GTAGCAGATATTTCATTTTCGACTATCGTATCGGTTTAAAGTGCAGAAATATGAGTGGAT 391271
Qy 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123
Db 391272 TGTTTACGTATCAATCCTGGCAACATTTGCTGAGATCGCGTCCGTCGCTGTTGTGAT 391331
Qy 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143
Db 391332 TGTGCGCGAGACAAAATATTTCGATTCGATTGTTGTTAAATGCAGGCTCTTTAGAAAAA 391391
Qy 144 AspLeuGlnGluLysTyrglyGluProThrProGlnAlaLeuLeuGluSerAlaMetArg 163
Db 391392 GATTTGCAAGAAAAATATGCGCAACCAACGCGAGAGCCCTGTTAGAAATCCGCAATTCGCT 391451
Qy 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183
Db 391452 CATGTAGAAATTTCTAGATCGTCTTAACATCGATCTTAAAGTGAAGCCTAAAGCCTCC 391511
Qy 184 AspValPheLeuAlaValIleGlySerTyrglyLeuLeuAlaLysGlnIleAspGlnProLeu 203
Db 391512 GATGTATTCTTAGCGGTTGAATCTTATCGTTTACTGGCTTAAGCAATTAACAGCCTTTA 391571
Qy 204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGly 223
Db 391572 CATTTAGGCATTTACAGAAGCAGGTGCGCACGCGTGGTGCAGTAAATCTCGAGTGGGT 391631
Qy 224 LeuGlyLeuLeuSerGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243
Db 391632 TTAGGAATGTTATTAGCTGAGGGCATTCGCGATACACACGCGCTCTCTTTGCGGCAGAT 391691
Qy 244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263
Db 391692 CTTGTAGAGAAATCAAAAGTCGTTTGTATATTTTGAATCTTTACGGATTCGTTCAAGA 391751
Qy 264 GlyIleAsnPhelIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283
Db 391752 GGAATTAACCTTTATTGCTTGCCCAACCTGTTCTCGCAAGAAATTTGATGTAATCGGTACA 391811
Qy 284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303
```

161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPhesGlnPheLysValSerVal 180
655 GCAATCGCATGTTGATATCTTGGACAGCGTGAATTCAGTCAAGTCAAGTGT 714
181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnLeuAsp 200
715 AAAGCGTCGATGCTTCTTGGCGTGGCTCTTATCGTTTATGGCGCAAAATTTGAT 774
201 GlnProLeuHisLeuGlyLeuThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220
775 CAACCACTTCCATCGGTATTACAGAAGCGGTGGGCTCGTTCTGTTCAAGTGAATCA 834
221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
835 GCAATTTGGTCTTGGTATGTTGGCTGAAGGTATCGCGATACGTTACGTATCTCAC 894
241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
895 GCGGCAGATCTCTGAGGAGTGAAGTGGTGGTTTATGATTTCTAAATCGTTACGATC 954
261 ArgSerArgGlyIleAsnPhelIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
955 CGCTCAGTGGCATCAACTTATTTGCTTGGCCCAACCTGTTACGCCAAGAATTTGATGT 1014
281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
1015 ATTGGTACGGTAAATGTTTGGAGCAGCGCTCGAAGATATTATCAGCGCGATGATGC 1074
301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
1075 TCTATTATTGTTGTAGTGAATGCCCGGTGAAGCCGAGGTTTCTACTTTAGTGTG 1134
321 ThrGlyCysAsnLysLysSerGlyLeuTyrGluAspGlyValArg--LysAspArgLeu 339
1135 GCTGGCGCGCAAAACCAAAAGTGGTTCTATGAAGATGGCTTCGCAAAAGAGCGTTT 1194
340 AspAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeu 359
1195 GATAATGACAAATATTATGATCAGCTTGAGCGCAAAATTCGCGCAAAAGCAGCAATGCTT 1254
360 AspGluAlaArgIleAspValGlnGlnValGlu 371
1255 GATGAAT-AACGTTATAAGATAAACCAAGTCGA 1289

RESULT 4
US-09-170-187-4
Sequence 4, Application US/09170187
Patent No. 6383745
GENERAL INFORMATION:
APPLICANT: Rather, Philip N.
TITLE OF INVENTION: Methods For Screening For Antimicrobials
TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlin & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/170,187
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/827,190
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1345 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 175..1272
OTHER INFORMATION: /gene= "aarC"
OTHER INFORMATION: /note= "Similar to E. coli GcpE protein listed by GenBank."
US-09-170-187-4

Alignment Scores:
Pred. No.: 1,02e-177 Length: 1345
Score: 1624.50 Matches: 325
Percent Similarity: 94.09% Conservative: 25
Best Local Similarity: 87.37% Mismatches: 21
Query Match: 87.06% Indels: 2
DB: 4 Gaps: 1

US-09-921-992-50 (1-372) x US-09-170-187-4 (1-1345)

Qy 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
Db 175 ATGCAATATGAATCACCGATAAAGACGTAATCCACCCGAAATTTATGTAGGTAAAGTG 234
Qy 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 235 CCTATTGGCGATGGTCTCCCATTTGCTGTCATCTATGACGAATACGCGACGCGAT 294
Qy 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
Db 295 GTTGAAGCCACTGTGCGGCAAAATCCCAATCATTGACGCTGTAGGTGTTGATATCGTCCGC 354
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
Db 355 GTGCTCTGTTCTACGATGGATGCGAGCAAGCGCTTTAAATTAATTAAAGCAGCGCTGAAT 414
Qy 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db 415 GTGCCATTGGTTGGGATATTTCATTGACTACCGTATCCGATGAAAGTGGCTGAATAT 474
Qy 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
Db 475 GGTGTTGACTGCTACGAATTAACCCAGGTAATATCGGCAGTGAAGAGCGTATTTCGCCAA 534
Qy 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
Db 535 GTGCTTGATGCTGCTCGTCATCACACATCTCTATCCGATAGGGTCAATGCGGCTCA 594
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
Db 595 CTGGAATAAGATATCCAGAAAAATACGCTGAGCAACACCTGAACGATTGGTTGAATCA 654
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPhesGlnPheLysValSerVal 180
Db 655 GCAATCGCATGTTGATATCTTGGACAGCGTGAATTCAGTCAAGTCAAGTGT 714
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 715 AAAGCGTCGATGCTTCTTGGCGTGGCTCTTATCGTTTATGGCGCAAAATTTGAT 774
Qy 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220

Db 121 GTCTGAAGCAACGGCTCAATCAATCAAGCGCTGGAACGGCTTGCGCTGATATCGTCCGT 180
Qy 61 ValSerValProThrMetAspAlaGluAlaPheLysLeuLeuLysGlnGlnValAsn 80
Db 181 GTATCCGTACCGACGATGGACGGCGGAGAGCGTTCAAACTCATCAACAGAGGTTAAC 240
Qy 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db 241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTCGCTGGAAGTAGGCGAATAC 300
Qy 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgMet 120
Db 301 GGCGTCCGATTTGCGGTATTAACCTCGCAATATCGTAATCGTAATGAAGAGCGTATTCGCATG 360
Qy 121 ValValAspCysAlaArgPheLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
Db 361 GTGGTTGACTGTGCGGCGGATAAACATTCGATCCGATTCGCTTAAACGCGGATCG 420
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyLupProThrProGlnAlaLeuLeuGluSer 160
Db 421 CTGGAAAAGATCTGCAAGAAAGTATGGCAACCGACCGCGGCGGTGCTGGATCT 480
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
Db 481 GCCATCGCTCATGTCATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 541 AAAGCCTCTGACGCTCTTCCTCGCTGTGAGTCTTATCGTTGTCGCAAAACAGATCGAT 600
Qy 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220
Db 601 CAGCCGTTGCACTTGGGGATCACCGAAGCGGFGTGGCGACGGCGGCGAATCAATCC 660
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
Db 661 GCCATTGGTTAGTCTGCTGCTGTCTGAAGGCATCGCGACACGCTGCGCGTATCGCTG 720
Qy 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 721 GCGGCGGATCCGGTCGAAGAGATCAAAAGTCGGTTTCGATATTTGAAATCGCTGCGTATC 780
Qy 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
Db 781 CGTTCGCGAGGATCACTTCATCGCTGCCGACCTGTTCGGCTCAGGAATTTGATGT 840
Qy 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
Db 841 ATCGGTACGGTTAAACCGCTGGAGCAACGCTTGAAGATATCATCACTCCGATGGACGTT 900
Qy 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
Db 901 TCGATTATCGCTGCTGTGTGTAATGCCCGCAGGTGAGCGCTGTTTCTACACTCGCGCTC 960
Qy 321 ThrGlyLysAsnLysSerGlyLeuTyrGluAspGlyValArgLys 336
Db 961 ACCGGCGCAACAAAGAAAGCGCCCTCTATGAAGATGGCGTGCGCAA 1008

RESULT 3

US-08-827-190-4
; Sequence 4, Application US/08827190
; Patent No. 5858367
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medien & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California

COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/827,190
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1345 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 175..1272
OTHER INFORMATION: /gene= "aarC"
OTHER INFORMATION: /note= "Similar to E. coli GcpE protein listed by GenBank."
US-08-827-190-4

Alignment Scores:
Pred. No.: 1,028-177 Length: 1345
Score: 1624.50 Matches: 325
Percent Similarity: 94.09% Conservative: 25
Best Local Similarity: 87.37% Mismatches: 21
Query Match: 87.06% Indels: 2
DB: 2 Gaps: 1

US-09-921-992-50 (1-372) x US-08-827-190-4 (1-1345)

Qy 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
Db 175 ATGCATATGATATCACCAGTAAAGAGCGTAATCCACCCGATTTATGTAGTAACTG 234
Qy 21 ProfileGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 235 CCTATTGGCGATGGTGTCTCCCATTCGTCTCAATCTATGACGAATACGCGCAGCGAT 294
Qy 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
Db 295 GTTGAAGCCACTGTGCGGCAATCCCAATCCTAGCGGTAGGTGTTGATATCGTCCG 354
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuLysGlnGlnValAsn 80
Db 355 GTGCTCTCTTACCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 414
Qy 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db 415 GTGCCATTGGTTGCGGATATTCACCTTGACTACCGTATCGCGATGAAAGTGGCTGAAT 474
Qy 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgMet 120
Db 475 GGTGTGTGATGCTGCTACGAAATTAACCCAGGATATATCGGAGTGAAGAGCGTATTCGCCAA 534
Qy 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
Db 535 GTCGTTGATGCTGCTGCTCATCACACATTCCTATCCGTATAGGGGTCAATGGCGGTCA 594
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
Db 595 CTGGAAAAGATATCAAGAAAAATACGCTGAGCCAAACACCTGAAGCATTTGTTGAATCA 654


```
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-827-190-8

Alignment Scores:
Pred. No.: 1,43e-185 Length: 1010
Score: 1691.00 Matches: 336
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.62% Indels: 0
DB: 2 Gaps: 0

US-09-921-992-50 (1-372) x US-08-827-190-8 (1-1010)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
DB 1 ATGCATAACAGGCTCCAAATTCACGTAGAAAATCAACAGTATTACGTGGGAATGTG 60
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB 61 CCGATTGGCGATGGTCTCCATTCGCGCTACGTACCTGATGACCAATACGCGTACGACAGAC 120
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
DB 121 GTCGAAGCAACGGTCAATCAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGTCGT 180
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnInValAsn 80
DB 181 GTATCCGTACCGACGATGGACGGCGGAGAGCGTTCAAACTCATCAACACAGAGGTTAAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
DB 241 GTCCCGCTGTGCTGACATCCACTTCGACTATCGCATTCGCTGAAAGTAGGCGGAATAC 300
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluArgIleArgMet 120
DB 301 GCGCTCGATTGTCTGGGTATTAACCTCGCAATATCGGTAAATGAAGAGCGTATTCGCATG 360
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
DB 361 GTGGTTGACTGTCGCGCGATAAACATTCCTCGTATTCGCTATTGGCGTTAAACGCGGATCG 420
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyLupProThrProGlnAlaLeuLeuGluSer 160
DB 421 CTGGAATAAGATCTGAGAAAAGATATGGCAACCGCGCGGCGGCTTGGTGGATCT 480
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
DB 481 GCCATCGCTCATGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB 541 AAAGCGCTGACGCTCTCTCGCTGTTGAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValLysSer 220
DB 601 CAGCCGTTGATCTGGGGATCACCGAAGCCGGTGGTGGCGCGAGCGGGGCGAGTAAATCC 660
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
DB 661 GCCATTGGTTTAGTCTGCTGCTGCTGAAGGCATCGCGACACGCTGCGCGTATCGCTG 720
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB 721 GCGGCGCATCCGTCGAAGAGATCAAGTCGGTTTCGATATTTGAAATCCGCTGCGGTATC 780
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
DB 781 CGTTCGGAGGATCAACTTCATCGCTGCCACCTGTCGCCGTCAGGAATTGATGTT 840
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
DB 841 ATCGGTACGTTTAAACGCGCTGGAGCAACGCTGGAGATATCATCATCCGATGGACGTT 900
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
DB 901 TCGATTATCGCTGCTGCTGATGATGCCCGAGTGGCGCTGTTTCTACACTCGCGGTC 960
QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLys 336
DB 961 ACCGGCGCAACAAGAAAGCGCTCTATGAAGATGGCGTGGCAAA 1008

RESULT 2
US-09-170-187-8
; Sequence 8, Application US/09170187
; Patent No. 6383745
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,187
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/827,190
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1010 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-09-170-187-8

Alignment Scores:
Pred. No.: 1,43e-185 Length: 1010
Score: 1691.00 Matches: 336
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 90.62% Indels: 0
DB: 4 Gaps: 0

US-09-921-992-50 (1-372) x US-09-170-187-8 (1-1010)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
DB 1 ATGCATAACAGGCTCCAAATTCACGTAGAAAATCAACAGTATTACGTGGGAATGTG 60
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB 61 CCGATTGGCGATGGTCTCCATTCGCGCTACGTACCTGATGACCAATACGCGTACGACAGAC 120
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
DB 121 GTCGAAGCAACGGTCAATCAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGTCGT 180
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnInValAsn 80
DB 181 GTATCCGTACCGACGATGGACGGCGGAGAGCGTTCAAACTCATCAACACAGAGGTTAAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
DB 241 GTCCCGCTGTGCTGACATCCACTTCGACTATCGCATTCGCTGAAAGTAGGCGGAATAC 300
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluArgIleArgMet 120
DB 301 GCGCTCGATTGTCTGGGTATTAACCTCGCAATATCGGTAAATGAAGAGCGTATTCGCATG 360
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
DB 361 GTGGTTGACTGTCGCGCGATAAACATTCCTCGTATTCGCTATTGGCGTTAAACGCGGATCG 420
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyLupProThrProGlnAlaLeuLeuGluSer 160
DB 421 CTGGAATAAGATCTGAGAAAAGATATGGCAACCGCGCGGCGGCTTGGTGGATCT 480
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
DB 481 GCCATCGCTCATGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB 541 AAAGCGCTGACGCTCTCTCGCTGTTGAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValLysSer 220
DB 601 CAGCCGTTGATCTGGGGATCACCGAAGCCGGTGGTGGCGCGAGCGGGGCGAGTAAATCC 660
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
DB 661 GCCATTGGTTTAGTCTGCTGCTGCTGAAGGCATCGCGACACGCTGCGCGTATCGCTG 720
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB 721 GCGGCGCATCCGTCGAAGAGATCAAGTCGGTTTCGATATTTGAAATCCGCTGCGGTATC 780
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
DB 781 CGTTCGGAGGATCAACTTCATCGCTGCCACCTGTCGCCGTCAGGAATTGATGTT 840
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
DB 841 ATCGGTACGTTTAAACGCGCTGGAGCAACGCTGGAGATATCATCATCCGATGGACGTT 900
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
DB 901 TCGATTATCGCTGCTGCTGATGATGCCCGAGTGGCGCTGTTTCTACACTCGCGGTC 960
QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLys 336
DB 961 ACCGGCGCAACAAGAAAGCGCTCTATGAAGATGGCGTGGCAAA 1008
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 23, 2003, 17:21:14 ; Search time 76 Seconds
(without alignments)
2160.454 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866

Sequence: 1 MHNQAPIQRKSTRIYGVN.....RAKASQLEARRIDVQOQVEK 372

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US09921992/runat_21112003_162547_18433/app_query.fasta_1.519
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09921992@cgn_1_56_orunat_21112003_162547_18433 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1691	90.6	1010	2	US-08-927-190-8
2	1691	90.6	1010	4	Sequence 8, Appli
3	1624.5	87.1	1345	2	US-08-927-190-8
4	1624.5	87.1	1345	4	Sequence 4, Appli
5	1603.5	85.9	1830121	4	US-09-170-187-4
6	1603.5	85.9	1830121	4	US-09-557-884-1
7	1593.5	85.4	1079	2	US-09-643-990A-1
8	1593.5	85.4	1079	4	Sequence 1, Appli
9	1372.5	73.6	1644	4	US-08-927-190-7
10	1372.5	73.6	1644	4	Sequence 7, Appli
11	1363.5	73.1	2088	4	US-09-252-991A-6420
12	1363.5	73.1	2088	4	Sequence 6420, Ap
					Sequence 6131, Ap
					Sequence 6339, Ap
					Sequence 10, Appli

C 13	1353	72.5	886	4	US-09-170-187-10	Sequence 10, Appli
C 14	1195	64.0	1137	4	US-09-328-352-3780	Sequence 3780, Ap
C 15	812	43.5	886	2	US-08-827-190-9	Sequence 9, Appli
C 16	812	43.5	886	4	US-09-170-187-9	Sequence 9, Appli
C 17	764	40.9	555	4	US-09-252-991A-6251	Sequence 6251, Ap
C 18	751.5	40.3	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 19	751.5	40.3	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 20	607.5	32.6	543	4	US-09-252-991A-6209	Sequence 6209, Ap
C 21	470.5	25.2	1230025	4	US-09-198-452A-1	Sequence 1, Appli
C 22	288.5	15.5	1083	4	US-09-221-017B-309	Sequence 309, App
C 23	112.5	6.0	3177	4	US-09-252-991A-12534	Sequence 12534, A
C 24	111.5	6.0	3998	3	US-09-066-046-5	Sequence 5, Appli
C 25	111	5.9	1665	4	US-09-221-017B-1017	Sequence 1017, Ap
C 26	106.5	5.7	1854	4	US-09-252-991A-14350	Sequence 14350, A
C 27	106.5	5.7	2943	4	US-09-252-991A-14267	Sequence 14267, A
C 28	106	5.7	2259	4	US-09-252-991A-784	Sequence 784, App
C 29	106	5.7	1664976	4	US-08-916-421B-1	Sequence 1, Appli
C 30	105.5	5.7	2129	3	US-08-975-762-39	Sequence 39, Appli
C 31	105.5	5.7	2129	3	US-09-295-028-39	Sequence 39, Appli
C 32	105.5	5.7	2129	4	US-09-106-582-39	Sequence 39, Appli
C 33	103.5	5.5	2214	1	US-07-985-458-1	Sequence 1, Appli
C 34	101.5	5.4	1694	3	US-09-136-421-11	Sequence 11, Appli
C 35	101.5	5.4	2989	1	US-07-927-851-1	Sequence 1, Appli
C 36	101.5	5.4	2989	1	US-08-453-323-1	Sequence 1, Appli
C 37	101.5	5.4	2989	1	US-08-440-520-1	Sequence 1, Appli
C 38	101	5.4	1149	4	US-09-252-991A-13125	Sequence 13125, A
C 39	101	5.4	1398	4	US-09-252-991A-12816	Sequence 12816, A
C 40	101	5.4	1572	4	US-09-612-964-1	Sequence 1, Appli
C 41	100.5	5.4	1863	4	US-09-328-352-3604	Sequence 3604, Ap
C 42	100.5	5.4	2010	3	US-09-039-859-1	Sequence 1, Appli
C 43	100.5	5.4	11601	2	US-08-222-617A-3	Sequence 3, Appli
C 44	100.5	5.4	11601	2	US-08-222-617A-24	Sequence 24, Appli
C 45	99.5	5.3	3414	4	US-09-252-991A-8258	Sequence 8258, Ap

ALIGNMENTS

RESULT 1
US-08-827-190-8
; Sequence 8, Application US/08827190
; Patent No. 5858367
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aaC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/827,190
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,937
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1010 base pairs
; TYPE: nucleic acid

COMMENT

Contact: Ross Whetten
Forest Biotechnology Group
North Carolina State University
Dept. of Forestry, NC State University, 6113 Jordan Hall, Raleigh
NC, 27695-8008
Tel: 919-515-7800
Fax: 919-515-7801
Email: rosswhet@unity.ncsu.edu
Seq primer: 5' lambda Triplex2 Sequencing Primer.

FEATURES

Location/Qualifiers

1..601

/organism="Pinus taeda"

/mol_type="mRNA"

/db_xref="taxon:3352"

/clone="ST23E04"

/lab_host="E. coli BM25.8"

/clone_lib="Pine Triplex shoot tip library"

/notes="Organ: shoot tips; Vector: Lambda Triplex; Site 1: SfiI (A); Site 2: SfiI (B); Shoot tips (approx. 2 cm from apex) were collected during the spring, frozen and used for mRNA isolation. The SMART-PCR method (Clontech) was used to prepare a library from 1 ug total RNA, using the Lambda Triplex vector. Plasmid subclones in pTriplex were recovered by cre-lox excision in E. coli strain BM25.8 and sequenced from the 5' end."

BASE COUNT 162 a 103 c 166 g 165 t 5 others

ORIGIN

Alignment Scores:

Pred. No.: 7,76e-20 Length: 601
Score: 262.00 Matches: 57
Percent Similarity: 62.30% Conservative: 19
Best Local Similarity: 46.72% Mismatches: 40
Query Match: 14.04% Indels: 6
DB: 9 Gaps: 2

US-09-921-992-50 (1-372) x AW042702 (1-601)

QY 132 IleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrglyGlu 151
Db 24 ATCGGAATTGGCACAACCATGGAAGTCTTCCGATCGTACTATGATGATGATGAT 83
QY 152 ProThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeu 171
Db 84 ---TCTCCAGGGGTATGTTGGATCAGCATTTGAATTTCCAGCATTTGCCGAGATTG 140
QY 172 AsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSer 191
Db 141 GGTTTTCATAATTTGTGTTTTCATGAAGCGAGCGATCCTGTAGTCATGTTTCAGGCA 200
QY 192 TyrArgLeuLeuAlaLysGlnIle-----AspGlnProLeuHisLeuGly 206
Db 201 TACCGTTTACTTGTTCGGAGATGATGTGCAAGGATGGGATTTATCCATTGCAATTAGGA 260
QY 207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
Db 261 GTTACTGAAGTGTGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 320
QY 227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
Db 321 CTTTTCAGGATGTTTGGGTGATACTATTTCGAGTTTCCCTTACAGAACCTCCAGAGAG 380
QY 247 Gluile 248
Db 381 GAGATC 386

RESULT 15

BJ551332

LOCUS

BJ551332 603 bp mRNA linear EST 14-NOV-2002

DEFINITION BJ551332 K. Sato unpublished cDNA library, strain H602 adult,

cDNA clone bAh59b20 5', mRNA sequence.

ACCESSION

BJ551332

VERSION

BJ551332.1 GI:24969783

KEYWORDS

EST.

SOURCE

ORGANISM

Hordeum vulgare subsp. spontaneum
Hordeum vulgare subsp. spontaneum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.

REFERENCE

1 (bases 1 to 603)

AUTHORS

Sato, K., Saisho, D. and Takeda, K.

TITLE

Barley EST sequencing project in NIG and Okayama Univ

JOURNAL

Unpublished

COMMENT

Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp.

FEATURES

source

1..603

/organism="Hordeum vulgare subsp. spontaneum"

/mol_type="mRNA"

/strain="H602"

/db_xref="taxon:77009"

/clone="bah59b20"

/tissue_type="top three leaves"

/dev_stage="adult, heading stage"

/clone_lib="K. Sato unpublished cDNA library, strain H602

adult, heading stage top three leaves"

BASE COUNT 156 a 112 c 169 g 165 t 1 others

ORIGIN

Alignment Scores:

Pred. No.: 1.02e-19 Length: 603
Score: 261.00 Matches: 56
Percent Similarity: 63.93% Conservative: 22
Best Local Similarity: 45.90% Mismatches: 38
Query Match: 13.99% Indels: 6
DB: 12 Gaps: 2

US-09-921-992-50 (1-372) x BJ551332 (1-603)

QY 132 IleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrglyGlu 151
Db 42 CTGCGTATCGGACAAATCATGCTAGTCTTTCTGACCGGATATGAGTACTATGATGAT 101
QY 152 ProThrProGlnAlaLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeu 171
Db 102 ---TCTCCAGGGGAATGGTTCAGTCTGCTTTGGAAATTTGCTAGGATCTGTCCGAATTTG 158
QY 172 AsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSer 191
Db 159 GACTTCCATAACTTTGTATTTTCAATGAAGCAAGTAACCCCTGTTGTCATGTTCCAAAGCA 218
QY 192 TyrArgLeuLeuAlaLysGlnIle-----AspGlnProLeuHisLeuGly 206
Db 219 TATCGCTCTTGTAGCGGAATGTATTAACCTTGGATGGGATTTATCTTTGACATTTGGGA 278
QY 207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
Db 279 GTTACCGAAGCTGGTTCAGGGTGAAGATGGAGGATGAAGTCTGCTATTGCGCATCGGAACA 338
QY 227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
Db 339 CTTTTCAGGATGCTTGGGTGATACATCCGTGTATCCCTTCACAGAACCCACAGAGGAA 398
QY 247 Gluile 248
Db 399 GAAAT 404

Search completed: November 23, 2003, 19:28:18

Job time : 2608 secs

QY 90 AsPTyArgIleAlaLeuLysValAlaGluTyArgValAspCysLeuArgIleAsnPro 109
 DB 13 GACAGCGAGCCAGTTGAGCAATTTAGAGTACACAGAGATGACTATCAGAAGAACTC 72
 QY 110 GlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLysAsn 129
 DB 73 GAGCATATT-----GAGGAGGTTTTCACACCATTTGTCACCAATTTGTTGAAATGTAAGAAGTATGGA 126
 QY 130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGluLysTyr 149
 DB 127 CGTGCATATGCGAATGGGACAAACCATCGAGGCTTTTCAGATCGCATATGAGCTATTAT 186
 QY 150 GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAsp 169
 DB 187 GGGGAC---TCGCCTAGGGGAATGGTAGAATCAGCATTTTCAGTTTCAAGAATTTGTAGA 243
 QY 170 ArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaVal 189
 DB 244 AAGTTGGACTTTTCACAAATTTCTGCTCTCAATGAAAGCTAGCAATCCAGTAGTATGGTT 303
 QY 190 GluSerTyArgLeuLeuAlaLysGlnIle-----AspGlnProLeuHis 204
 DB 304 CAGGCGTATGCGCTTTCTGAGCTGAGATGTATGTCAGGATGGGACTATCCTTTACAC 363
 QY 205 LeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224
 DB 364 TTGGGAGTACTGAAGCTGGTGAAGTGAGGATGGAAGATGAGTCTGCAATCGGTATT 423
 QY 225 GlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspPro 244
 DB 424 GGCACACTTCTTCAGGATGTTTGGGTGATACAAATACGCGGTTTCCCTAACTGAAGCTCCA 483
 QY 245 ValGluGluIle 248
 DB 484 GAAGAGGAGATA 495
 RESULT 13
 ID AU186794 standard; RNA; EST; 489 BP.
 AC AU186794;
 SV AU186794.1
 XX 23-MAY-2003 (Rel. 75, Created)
 DT 23-MAY-2003 (Rel. 75, Last updated, Version 1)
 XX Porphyra yezoensis cDNA, clone:PF004b08_r, 5'end.
 DE EST (expressed sequence tag).
 KW Porphyra yezoensis
 XX Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.
 [1]
 RN Asamizu E., Nakamura Y.;
 RA Submitted (15-JUN-2001) to the EMBL/GenBank/DBJ databases.
 RL Erika Asamizu, Kazusa DNA Research Institute, The First Laboratory for
 RL Plant Gene Research; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 RL (E-mail:asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/,
 RL Tel:81-438-52-3935(ex.2336), Fax:81-438-52-3934)
 XX [2]
 RN Asamizu E., Nakajima M., Kitade Y., Saga N., Nakamura Y., Tabata S.;
 RA "COMPARISON OF RNA EXPRESSION PROFILES BETWEEN THE TWO GENERATIONS OF
 RT PORPHYRA YEZOENSIS (RHODOPHYTA)", BASED ON EXPRESSED SEQUENCE TAG FREQUENCY
 RT ANALYSIS";
 RL J. Phycol. 0:0-0(2003).
 XX J. Phycol. 0:0-0(2003).
 FH Key Location/Qualifiers

source 1..489
 /db_xref="taxon:2788"
 /mol_type="mRNA"
 /organism="Porphyra yezoensis"
 /clone="PF004b08_r"
 /dev_stage="sporophytes"
 /strain="TU-1"
 XX Sequence 489 BP; 100 A; 136 C; 170 G; 83 T; 0 other;
 Alignment Scores:
 Pred. No.: 3,42e-20 Length: 489
 Score: 264.00 Matches: 68
 Percent Similarity: 57.14% Conservative: 24
 Best Local Similarity: 42.24% Mismatches: 48
 Query Match: 14.15% Indels: 22
 DB: 6 Gaps: 5
 US-09-921-992-50 (1-372) x AU186794 (1-489)
 QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAsp 123
 DB 14 TGTTTGAGA---AGCCCCAAGAGAGCGGCGGAGTACTCGGACGAGGAGTTCAACCGGC 70
 QY 124 -----CysAlaArgAspLysAsn----- 129
 DB 71 TCAAGACAGCTCAAGGCCACCCTCACACCACTGGTGTTCGCTCAAGAGCAGCGGCA 130
 QY 130 IleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGluLysTyr 149
 DB 131 AGGCCATG-CGTATTGGGTCAACACACGGTCCCTGGCGGAGCGCATGATGTTACCTAC 189
 QY 150 GlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAsp 169
 DB 190 GGGGAC---ACCCCGCGGGGATGTCGAGTCCGGATGAGTGCAATTCAGATCTGCCCG 246
 QY 170 ArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaVal 189
 DB 247 GAGCTAGACTTTCACAACTGGTCTATTTCGATGAGGCGTCCCAAGTGCCAGTCATGATC 306
 QY 190 GluSerTyArgLeuLeuAlaLysGlnIle-----AspGlnProLeuHis 204
 DB 307 GCAGCTACCGCCAGCTGGCGCTCGGCTCGACGCTCGAGGCTGAGGGGTACGACTACCCATCCAC 366
 QY 205 LeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeu 224
 DB 367 CTGGGTGTGACTGAGCGCGGCGGAGTATGGCGGATGTCGCGGCTGGAAGAGCTCCATCGGTATC 426
 QY 225 GlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspPro 244
 DB 427 GGCACGCTCTTGGGAGGGGATCGGCGACACGCTGCGGGTGAGCTTGACAGAGAACCC 486
 QY 245 Val 245
 DB 487 ATG 489
 RESULT 14
 ID AU042702
 LOCUS
 DEFINITION ST23E04 Pine Triplex shoot tip library Pinus taeda cDNA clone
 ST23E04, mRNA sequence.
 ACCESSION AU042702
 VERSION AU042702.1 GI:5903147
 KEYWORDS EST.
 SOURCE Pinus taeda (loblolly pine)
 ORGANISM Pinus taeda
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
 REFERENCE 1 (bases 1 to 601)
 AUTHORS Whetten,R.W., Kinlaw,C.S., Retzel,E. and Sederoff,R.R.
 TITLE The Pine Gene Discovery Project
 JOURNAL Unpublished

QY 48 IleLeuAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAsp 67
 Db 358 GTTATGAGTACACGATTAAGGGCTGATTTGTTAGATAACATCCAGGCTAGAAAG 417
 QY 68 AlaAlaGluAlaPhe-----LysLeuIleLysGlnValAsnValPro 82
 Db 418 GAAGCTGATGCTGCTTTGAGATTAAAGAACACTCTTTGTTCAAGAGAAATTACAAACATCCCC 477
 QY 83 LeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyVal 102
 Db 478 CTAGTGGCTGATATTCATTTTGGCCGACAGATGCTTTAGAGTGGCTGAATGC---TTT 534
 QY 103 AspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlu----- 115
 Db 535 GACAAATTCGTCTCAACCCAGGGAATTTGCTGATCGCGTCCCAATTTGAGCAGCTT 594
 QY 116 -----GluArgIleArgMetVal----- 121
 Db 595 GAATATACTGAAGATGATTATCAAAAGAGCTTTGAGCATATCAGAAAGTCTTCTCCCG 654
 QY 122 ---ValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
 Db 655 TTGTTGAGAAATCGAAGCAGTATGGAAGCAATCGGTATAGAACAAATCATGGAAT 714
 QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
 Db 715 CTGCTGACCGCATATGATGTTACTATGTTGAT---TCTCCAGCGGAATGTTGAGTCT 771
 QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
 Db 772 GCTTTGGAATTTTCAGGATCTGTGCG- AAGCTGGACTTCCATAACTTTGTTTTCATG 830
 QY 181 LysAlaSerAsp 184
 Db 831 AAGCAAGTAAC 842

RESULT 10

CA022320
 LOCUS HZ42M1r HZ Hordeum vulgare subsp. vulgare cDNA clone HZ42M1
 DEFINITION S-PRIME, mRNA sequence.

ACCESSION CA022320

VERSION CA022320.1 GI:24299694

KEYWORDS EST.

SOURCE Hordeum vulgare subsp. vulgare

ORGANISM Hordeum vulgare subsp. vulgare

REFERENCE Ruckluch V., Zhang H., Weschke, W., Fotokina, E. and Wobus, U.
 Barley ESTs from developing seeds

AUTHORS Ruckluch V., Zhang H., Weschke, W., Fotokina, E. and Wobus, U.

TITLE Barley ESTs from developing seeds

JOURNAL Unpublished

COMMENT Contact: Stein Nils

Molecular Markers Group, Department Genbank

Institute of Plant Genetics and Crop Plant Research (IPK)

Corrensstr. 3, 06466, Gatersleben, Germany

Tel: 039482-5522

Fax: 039482-5595

Email: stein@ipk-gatersleben.de

Insert length: 602 Std Error: 0.00

Plate: 42 row: M column: 11

Seq primer: M13rev.

Location/Qualifiers

1. 602

/organism="Hordeum vulgare subsp. vulgare"

/mol_type="mRNA"

/cultivar="barke"

/db_xref="GABI:275726"

/db_xref="taxon:112509"

/clone="HZ42M1"

/tissue_type="pericarp"

/dev_stage="0-7 DAP (days after pollination)"

RESULT 11

BJ481080

LOCUS

DEFINITION

CDNA clone bah58101 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

BJ481080 610 bp mRNA linear EST 23-MAY-2002

BJ481080 K. Sato unpublished cDNA library, strain H602 adult,

heading stage top three leaves Hordeum vulgare subsp. spontaneum

CDNA clone bah58101 5', mRNA sequence.

BJ481080

BJ481080.1 GI:21159548

EST.

/lab host="XL10-Gold"

/clone lib="HZ"

/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of

CDNA); Site 2: XhoI (3'-end of CDNA); pericarp 0-7

DAP(days after pollination). Due to a cloning artefact

caused by the kit, in most cases the EcoRI site is NOT

present, as well as the EcoRI adapter used for cloning. To

excise the insert, restriction sites upstream EcoRI should

be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the

cloning system used Blue/white selection for recombinats

is not 100% reliable. Average insert size is 900 bp"

BASE COUNT 169 a 110 c 156 g 167 t

ORIGIN

Alignment Scores:

Pred. No.: 7,14e-21 Length: 602

Score: 271.00 Matches: 70

Percent Similarity: 51.52% Conservative: 32

Best Local Similarity: 35.35% Mismatches: 62

Query Match: 14.52% Indels: 34

DB: 14 Gaps: 6

US-09-921-992-50 (1-372) x CA022320 (1-602)

QY 55 GlyAlaAspIleValArgValSerValProThrMetAspAlaAlaGluAlaPhe----- 72

Db 15 GCGCTGATTTTGTAGATAACCGTCCAGGGTAAAAAGAGAGCTGATCGCTTGTAG 74

QY 73 -----LysLeuIleLysGlnValAsnValProLeuValAlaAspIleHisPhe 89

Db 75 ATTAAGAACACTCTTGTCCAGAGAAATTACAACATCCCTCTAGCGGCCGATATCCATTTT 134

QY 90 AspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnPro 109

Db 135 GCGCTACAGTACGTTTAAAGTGGCTGATGC---TTTGACAAAATCGTGTAAACCA 191

QY 110 GlyAsnIleGlyAsnGlu----- 115

Db 192 GGAATCTTTGCGATCGCGTCCCAATTTGAAAGTGAATATATCTGAAGACCGATTAC 251

QY 116 -----GluArgIleArgMetVal-----ValAspCysAlaArgAsp 127

Db 252 GAAAGGAGCTTGAACACATTCAGAGGGTCTTTTCTCCATTCGTTGAGAAATGCAAGAAG 311

QY 128 LysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGlu 147

Db 312 TATGGAAGAGCCATCGGTATCGAACAAATCATGTAGTCTTTCTGACCGGATATGAGC 371

QY 148 LysTyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHis 167

Db 372 TACTATGGTGAT---TCTCAAGGGGAATGGTTGAGTCTGCTTTGGAATTTGCTAGGATC 428

QY 168 LeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeu 187

Db 429 TGTCCGAATTTTGGACTTCCATACTTTGATTTTCAATGAAAGCAAGTAACCTGTGTTC 488

QY 188 AlaValGluSerTyrArgLeuLeuAlaLysGlnIle-----AspGlnPro 202

Db 489 ATGGTCCAAGCATATCGCTGTTGAGCGGAAATGATATACTTGGATGGGATATCTCT 548

QY 203 LeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220

Db 549 TTGCACCTTGGAGTTACCGAAGCTGGTGGGAGGATGGGAGGATGAAGTCT 602

```

/db_xref="taxon:5855"
/dev_stage="asexual blood forms"
/lab_host="Gaimiri boliviensis"
/clone_lib="Pv MBN #30"
/notes="Vector: pBluescript SK(+) vector DNA, phagemid
excised from lambda ZAP; Site 1: EcoR V; Site 2: EcoR V;
Host leukocytes were extracted from P. vivax infected
blood using the following methods: first, infected blood
was activated by the addition of 0.5 ml of ADP (40mg/ml)
per 10 ml blood. Then blood was passed over a column of
acid washed 0.1 mm glass beads, then through a Plasmidipur
filter, followed by passage through a column of pre-wet
Whatman Cfil powder (1:2 ratio volume of blood to Cfil),
and finally centrifuged through a 50% Percoll density
cushion. Purified DNA was digested with mung bean nuclease
in the presence of 4% formamide at 50°C as described
(Vernick, K.D., ImberSKI, R.B., and McCutchan, T.F. 1988.
Nucleic Acids Research 16:6883-6896). Digested DNA was
blunt-ended using T4 DNA Polymerase and size fractionated
over a Sepharose CL-2B column. Fractions in the size range
500bp-4kb were ligated into the Eco RV site of pBluescript
SK(+), and E. coli XL-10 Gold transformed with the
ligation mixture."

```

BASE COUNT 178 a 150 c 119 g 206 t 5 others
ORIGIN

Alignment Scores:
Pred. No.: 5,01e-25 Length: 658
Score: 307.50 Matches: 79
Percent Similarity: 53.85% Conservative: 33
Best Local Similarity: 37.98% Mismatches: 61
Query Match: 16.48% Indels: 35
DB: 28 Gaps: 6

US-09-921-992-50 (1-372) x AZ570993 (1-658)

```

QY 73 LysLeuIleLysGlnValAenValProLeuValAlaAspIleHisPheAspTyrArg 92
DB 642 AAAATGGAGCTCAAAATATTACATTCCTCGTTAGTAGACATACATTTATCCCAAA 583
QY 93 IleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnProGlyAenIle 112
DB 582 ATTCTCTCATGCAAGCAGATGTA---TTTGATAAAATACGCATACCCAGCGAATTAC 526
QY 113 GlyAenGluLuarG----- 117
DB 525 GTCGATGGGAGNAGAAATGGATCAATAAGATTACAAACGAGAGAGANTTCGATCAA 466
QY 118 -----IleArgMetValValAspCysAlaArgAspLys 128
DB 465 GGGAAATTATTTCAGNAGAGTTTCATCCCTTGATTGAAAGTGTAGAGG---TTA 409
QY 129 AenIleProIleArgIleGlyValAenAlaGlySerLeuGluLysAspLeuGlnGluLys 148
DB 408 AATAGAGCCATACGAATCGGCACAAATCACGGCTCCCTATCTTNTAGGCTACTCTCTTC 349
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 348 TATGGGGAT---ACCCNNTTAGTATGTCGATCGCGCTTTTCAATTTTCCCGATTGTGT 292
QY 169 AspArgLeuAsnPhaAspGlnPhelValSerValLysAlaSerAspValPheLeuAla 188
DB 291 GTACAGATAACTCTCTCAATGGTGTCTTCATGAGGCTTCCACCGCTATATATG 232
QY 189 ValGluSerTyrArgLeuLeu---AlaLysGlnIleAspGln----- 201
DB 231 ATACAGTCGTATAGGCTACTCGTGGCGAGGCAATATGAAAGGATGGCAACGGTTACTA 172
QY 202 ---ProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
DB 171 TTCCCACTGCATTTGGCGCTTACCGAAGCAGGGTTTGGGGATAACGGAGGATAAATCC 112
QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240

```

```

Db 111 TATTTGGGATAGGTTGCTGCTGTACGACGGGATAGGACACCATCAGGATTCGTTA 52
QY 241 AlaAlaAspProValGluGluIle 248
Db 51 ACTGAGGACCCCTGGGAGGAATTA 28

RESULT 9
LOCUS CB628479
DEFINITION OSIIEB04B06.f OSIIEB Oryza sativa (indica cultivar-group) cDNA
clone OSIIEB04B06 5', mRNA sequence.
ACCESSION CB628479
VERSION CB628479.1 GI:29623468
KEYWORDS EST.
SOURCE Oryza sativa (indica cultivar-group)
ORGANISM Oryza sativa (indica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 852)
AUTHORS Jantasuriyarat,C., Lu,G., Cowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
JOURNAL Unpublished
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210086, Tucson, AZ
85721-0086, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cga cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 04 row: B column: 06
Seq primer: gta aaa cga cga cca gtcg.
FEATURES
source
1..852
/organism="Oryza sativa (indica cultivar-group)"
/mol_type="mRNA"
/cultivar="IR36"
/db_xref="taxon:39946"
/clone="OSIIEB04B06"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSIIEB"
/notes="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; 24 hrs after inoculation with Rice Blast (P06-6-3)"
BASE COUNT 233 a 172 c 223 g 224 t
ORIGIN

Alignment Scores:
Pred. No.: 2.69e-21 Length: 852
Score: 276.50 Matches: 72
Percent Similarity: 53.43% Conservative: 37
Best Local Similarity: 35.29% Mismatches: 66
Query Match: 14.82% Indels: 30
DB: 14 Gaps: 5

US-09-921-992-50 (1-372) x CB628479 (1-852)
QY 8 GlnArgAspLysSerThrArgIleTyrValGlyAenValProIleGlyAspGlyAlaPro 27
DB 238 AGGAGGAGAAAAACCCGACCTGTGATGTTGGGATGTGCCACTTGGCAGTCATATCC 297
QY 28 IleAlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAenGln 47
DB 298 ATTAGGATTCAGATATGACCACCTCGGATACCAAGGATGTTGCTAAAAACCGTAGGAG 357

```



```

Db      71 TGC---TTTGACAAAATAGCTGTCAATCCTGGAAATCTTCTGCACAGCGGAGCCAGTTT 127
QY      115 -----GluGluArg 117
Db      128 GAGCAATTAGATACACAGAGATGACTATCAGAAAGAACTCGAGCATATTGAGGAGTT 187
QY      118 IleArgMetValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsn 137
Db      188 TTTACACCATGTGTGGAATAATGTAAGAGATGAGGAGCTGCATCGCATTTGGACAAAC 247
QY      138 AlaGlySerLeuGluLysAspLeuGlnGluLysTyrglyGluProThrProGlnAlaLeu 157
Db      248 CATGGAGGCCCTTCAGATCGCAATTATCAGCTATTATGGGAC---TCGCCATGGGAATG 304
QY      158 LeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLys 177
Db      305r GTAGATACAGCATTTGAGTTTGCAGAAATTTGTAGAAAGTTGGACTTTTCACAAATTCGTC 364
QY      178 ValSerValLysAlaSerAspValPheLeuAlaValGluSerTyrglyArgLeuLeuAlaLys 197
Db      365 TTTCTCAATGAAGTAGCAATCAGTAGTATGTTTTCAGCGGTATCGCTTCTCTAGCT 424
QY      198 GlnIle-----AspGlnProLeuHisLeuGlyIleThrGluAlaGlyGly 212
Db      425 GAGATGTATGTTTCAGGATGGGACTATCTTACACTTGGAGATTACTGAAGCTGCTGAA 484
QY      213 AlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuSerGluGlyIle 232
Db      485 GGTGAGGTGGAGGATGAAGTCTCAATCGGTATTGGCACACTTCTTCAGGATGTTTG 544
QY      233 GlyAspThrLeuArgValSerLeu 240
Db      545 GGTGATACATACGGGTTTCCCTA 568

```

```

RESULT 7
BH235010
LOCUS      BH235010
DEFINITION PH_05.x Spiroplasma kunkelii H Spiroplasma kunkelii genomic clone
ACCESSION PH_05.x, genomic survey sequence.
VERSION    BH235010
KEYWORDS   GSS.
SOURCE     Spiroplasma kunkelii
ORGANISM   Spiroplasma kunkelii
REFERENCE  Bacteria; Firmicutes; Mollicutes; Entomoplasmatiales;
            Spiroplasmataceae; Spiroplasma.
AUTHORS    Hogenhout, S.A.
TITLE      Genomic sequences from Spiroplasma kunkelii strain M2
JOURNAL    Unpublished
COMMENT    Contact: Hogenhout SA
            Department of Entomology
            The Ohio State University-OARDC
            120 Thorne Hall, 1680 Madison Avenue, Wooster, OH 44691, USA
            Tel: 330 263 3730
            Fax: 330 263 3686
            Email: hogenhout.1@osu.edu
            Class: HindIII fragments.
FEATURES   Location/Qualifiers
            1..720
            /organism="Spiroplasma kunkelii"
            /mol_type="genomic DNA"
            /strain="M2"
            /db_xref="taxon:47834"
            /clone="PH_05.x"
            /clone_lib="Spiroplasma kunkelii H"
BASE COUNT 255 a 96 c 132 g 233 t 4 others
ORIGIN
Alignment Scores: 1.32e-25 Length: 720
Pred. No.: 313.00 Matches: 67
Score:

```

```

Percent Similarity: 62.20% Conservative: 35
Best Local Similarity: 40.85% Mismatches: 61
Query Match: 16.77% Indels: 2
DB: 28 Gaps: 0
US-09-921-992-50 (1-372) x BH235010 (1-720)
QY      190 GluSerTyArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThrGlu 209
Db      31 CAAGCTTATACCTTAGCTAGTAAGAATTAATATCCCTTCATCTAGGGATGCT-GAA 89
QY      210 AlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuSer 229
Db      90 GCTGTAGTCATCATACCGAACAANNAATCATGTAGTGGTTGTGCACCGCTTCTTTT 149
QY      230 GluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIleLys 249
Db      150 AATGTTATTGGTGATACGATCCGAATTATGTTATCAACTGTATCCCAATCAGCAGGTGGAA 209
QY      250 ValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIleAla 269
Db      210 GTTGTAAACGAATGTTAAATTCATTAGGACTTATGATAACATTTGTCATATAATTGCT 269
QY      270 CysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGluGln 289
Db      270 TGCCCAACATGTGGTGGTTAGATATGACCTTTTCCCGCTTGTAAAGAAATAGAAGAA 329
QY      290 ArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsnGly 309
Db      330 TATACAAAGATTTTAAATTCATTTAAATTTGCCATTTTAGGTTTGTGTTGTTAATGTT 389
QY      310 -ProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGlyLe 329
Db      390 CCCAGAGAGACTAAACAGCTGATTAGGAGTTGCTGGTGTGTTAAATATGTTGGAATTAT 449
QY      329 uTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeuG 349
Db      450 TTTTAAAGAGTAAATTTATAAATTCATGTAACAAAGAAAGATCTTGTTCAGAAATATAA 509
QY      349 uAlaArgIle 352
Db      510 ACTATTAATT 519
RESULT 8
AZ570993/c
LOCUS      AZ570993
DEFINITION 281PVD07 Pv MEN #30 Plasmodium vivax genomic 3', genomic survey
ACCESSION AZ570993
VERSION    AZ570993
KEYWORDS   GSS.
SOURCE     Plasmodium vivax (malaria parasite P. vivax)
ORGANISM   Plasmodium vivax
REFERENCE  Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS    1 (bases 1 to 658)
            Carlton, J.M.-R. and Dame, J.B.
TITLE      The Plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL    Parasitol. Today (Regul. Ed.) 16 (10), 409 (2000)
COMMENT    Contact: Dame JB
            Dept. of Pathobiology, College of Veterinary Medicine
            University of Florida
            2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
            Tel: 352 392 4700
            Fax: 352 392 9704
            Email: damej@mail.vetmed.ufl.edu
            Seq primer: M13(-20) forward
            Class: shotgun.
FEATURES   Location/Qualifiers
            1..658
            /organism="Plasmodium vivax"
            /mol_type="genomic DNA"
            /strain="Salvador I (Collins, W. 1972. J. Parasitol. 69,
            497-598)"

```



```

QY      247 Gluile 248
Db      602 GAGATT 607

RESULT 4
AY104363
LOCUS   Zea mays PCO116062 mRNA linear HTC 16-OCT-2002
DEFINITION Zea mays PCO116062 mRNA sequence.
ACCESSION AY104363
VERSION   AY104363.1 GI:21207441
KEYWORDS HTC.
SOURCE   Zea mays
ORGANISM Zea mays
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
          clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1442)
AUTHORS   Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whittitt, M.S.,
          Arthur, L.W., Hanafey, M., Morgante, M. and Tingey, S.V.
TITLE     Maize Mapping Project/DuPont Consensus Sequences for Design of
          Overgo Probes
JOURNAL   Unpublished (2002)
REFERENCE 2 (bases 1 to 1442)
AUTHORS   Coe, E.H.
TITLE     Direct Submission
JOURNAL   Submitted (25-APR-2002) Maize Mapping Project, University of
          Missouri, Columbia, MO 65211, USA
COMMENT   If you are interested in getting corresponding physical clones,
          these are publicly available from ZmDB and may be found by BLAST
          searching at MSL, maizemap.org; ZmDB, www.zmdb.iastate.edu; TIGR,
          www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
          maize cDNA sequences is either Virginia Walbot, Stanford or Pat
          Schnable, Iowa State, then clones may be requested from ZmDB:
          www.zmdb.iastate.edu.
FEATURES             Location/Qualifiers
     source           1..1442
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /db_xref="MaizeDB:636753"
                     /db_xref="taxon:4577"
     clone            lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
     notes            /notes="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT  381 a 282 c 374 g 399 t 6 others
ORIGIN

Alignment Scores:
Pred. No.:      9.68e-27      Length:      1442
Score:          326.50      Matches:      99
Percent Similarity: 49.66%      Conservative: 46
Best Local Similarity: 33.90%      Mismatches:  77
Query Match:    17.50%      Indels:      71
DB:              11      Gaps:        11

US-09-921-992-50 (1-372) x AY104363 (1-1442)
QY      65 ThrMetAspAlaGluAlaPheLys-----LeuIleLysGlnValAsn 80
Db      3 ACAGGGAAGCTGATGCTGCTTTGAGATCAAGAACACTCTGGTTCAAGAAATTACAAC 62
QY      81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db      63 ATTCCACTAGTGGCCGATATTCATTTTCTCCACGCTAGCCCTAAGGGTGGCAGATGT 122
QY      101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlu----- 115
Db      123 ---TTTGACAAAATTCGTGTGACCCAGGAAATTTTCTGTATCTCGTCTCAATTGAA 179

```

```

QY      116 -----GluArgIleArgMetVal--- 121
Db      180 AAGCTGGAATATACTGACGACGACTACCAAAAGAGCTAGACATATTGAGAAGGTGTTT 239
QY      122 -----ValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAla 138
Db      240 TCTCCATTAGTTGAGAAATGCAAGCAGTAGTATGAAGAGCAATGCGAATAGGAACAATCAT 239
QY      139 GlySerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeu 158
Db      300 GGTAGTCTTTTCGACCGCATATAGCTACTATGGTGAT---TCTCCAGGGGAATGGTT 356
QY      159 GluSerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysVal 178
Db      357 GAGTCTGCTTGGAAATTCCTAGAAATATGTCGAAGTTGGACTTCCACAACATTGTATTT 416
QY      179 SerValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGln 198
Db      417 TCGATGAAAGCTAGTAACCCCTGTATCATGTGTCGAAGCATATCGCTGCTTGTGGCAGAA 476
QY      199 Ile-----AspGlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAla 213
Db      477 ATGTATTAACCTAGGATGGGATTTATCCCTTGCACCTGGGTGTACAGAGCTGGAGGGGT 536
QY      214 ArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGly 233
Db      537 GAAGATGA-AGGATGAAATCTGCTATTGGCATTTGGGACACTGCTAATGATGCTTTGGGT 595
QY      234 AspThrLeuArgValSerLeuAlaAlaAspProValGluGluLeu----- 248
Db      596 GATACAATCCGTCTCTCCCTCAGAACACCACAGAGAAGAGAGATTGATCCTTCCCAAAGG 655
QY      249 -----LysValGlyPheAspIleLeuLysSe 257
Db      656 TTGGCAAAATCTTGGGACGCGAGCGCCCAACCTTCAAAATTTGGGTGGCCCCCATTTGAAG-- 713
QY      257 rLeuArgIleArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnG 277
Db      714 -----AAAAGCACAGGGCTTATTTGATTTCACG----- 743
QY      277 uPheAspValIleGlyThrValAsnAlaLeu---GluGlnArgLeuGluAspIleIleTh 296
Db      744 -----GTAGGAGTGTCAATTCCTTTTCAGAGAGGGGTGAGGAAGTTGACTAC 793
QY      296 rProMetAspValSerIleleGlyCysValVal 307
Db      794 AGAAATG-----TCCTGCATCGTG 812

RESULT 5
BZ569685/c
LOCUS   BZ569685 1268 bp DNA linear GSS 17-DEC-2002
DEFINITION pac82-164_970.s1 pac82-164 Pseudomonas aeruginosa genomic clone
          pac82-164_970, genomic survey sequence.
ACCESSION BZ569685
VERSION   BZ569685.1 GI:27204646
KEYWORDS GSS.
SOURCE   Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
          Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
          Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1268)
AUTHORS   Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
          Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE     Whole-Genome-Sequence variation among multiple isolates of
          Pseudomonas aeruginosa library
JOURNAL   J. Bacteriol., (2002) In press
COMMENT   Contact: Chris K. Raymond
          Genome Center
          University of Washington
          Box 352145, Seattle, WA 98105-2145, USA
          Tel: 2062216954
          Fax: 2066857244
          Email: craymond@u.washington.edu

```

```

Qy 44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63
Db 6 ACCGTAGAGGAGTTATGAGGATAGCAGATAAAGGGGCTGATTTGTTAGATAACAGTC 65
Qy 64 ProThrMetAspAlaAlaGluAlaPhe-----LysLeuIleLysGlnGln 78
Db 66 CAGGGTAGAAGGAAGCTGATGCTGCTTTGAGATTAGAACAACACTCTTGTTCAGAAGAT 125
Qy 79 ValAsnValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAla 98
Db 126 TACAACAATCCCTAGTGGCTGATATTCATTTTCCCGGACAGCTGCTTTAAGAGTGGCT 185
Qy 99 GluTyrGlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlu----- 115
Db 186 GAATGC---TTTGACAAAATTCGTGTCAACCCAGGGAATTTTCTGATCGCCGTGCCAA 242
Qy 116 -----GluArgIleArgMet 120
Db 243 TTTGAGCAGCTTGAATATACTGAAGATGATTATCAAAAAGAGCTTGAGCATATCGAGAAG 302
Qy 121 Val-----ValAspCysAlaArgAspLysAsnIleProIleArgIleGlyVal 136
Db 303 GTCTTCTCCCGTTGGTTGAGAAATCAAGCAGCATGGAAGCAATGCGTATAGGAACA 362
Qy 137 AsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAla 156
Db 363 AATCATGGAAGTGTCTGACCCGCAATAGTACTATGTTGAT---TCTCCACGGGA 419
Qy 157 LeuLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPhe 176
Db 420 ATGGTTGAGTCTGCTTTGGAATTTGCCAGGATCTGTCCGAAGCTGGAACACTTCCATACT 479
Qy 177 LysValSerValLysValAspValPheLeuAlaValGluSerTyrArgLeuLeuAla 196
Db 480 GTGTTTCAATGAAGCAAGTACCTGTTATCATGTCGAACATATCGCTTGTGTA 539
Qy 197 LysGlnIle-----AspGlnProLeuHisLeuGlyIleThrGluAlaGly 211
Db 540 GCAGAAATCTATAACCTAGGCTGGGATATCTTTCACCTTGGAGGTACAGAAGCTGGA 599
Qy 212 GlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeuSerGluGly 231
Db 600 GAGGGTGAAGATGGAGGATGAAGTCTGCCATTGGCATTGGAACACTTCTGATGGATGC 659
Qy 232 IleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
Db 660 TTGGCGATACAAATCCGTGTCTCCCTCACGGAACCACTTGAAGAAGAGATT 710

RESULT 3
CB619369
LOCUS
DEFINITION
  OSIIeA03003.f OSIIeA Oryza sativa (indica cultivar-group) cDNA
  clone OSIIeA03003 5', mRNA sequence.
ACCESSION
  CB619369
VERSION
  CB619369.1 GI:29614356
KEYWORDS
  EST.
SOURCE
  Oryza sativa (indica cultivar-group)
  Oryza sativa (indica cultivar-group)
  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
  Ehrhartoideae; Oryzeae; Oryza.
  1 (bases 1 to 853)
REFERENCE
  Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
  Kudrna,B., Dean,R., Soderlund,C., Wing,R. and Wang,G.
  Large-scale identification of ESTs involved in the interaction
  between rice and Magnaporthe grisea
  Unpublished
  Contact: Rod Wing
  Arizona Genomics Institute
  University of Arizona
  Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
  85721-0088, USA
  Tel: 520 626 3967

```

```

FEATURES
  source
    1..853
    /organism="Oryza sativa (indica cultivar-group)"
    /mol_type="mRNA"
    /culturvar="IR36"
    /db_xref="taxon:39946"
    /clone="OSIIeA03003"
    /tissue_type="Leaf"
    /dev_stage="3 week"
    /lab_host="DH10B"
    /clone_lib="OSIIeA"
    /note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
    XhoI; Lesion Mimic SPL 11"
  BASE COUNT      222 a   168 c   223 g   240 t
  ORIGIN
    Alignment Scores:
    Pred. No.:      1,35e-29      Length:      853
    Score:          348.50        Matches:    81
    Percent Similarity: 56.93%    Conservative: 34
    Best Local Similarity: 40.10% Mismatches:    58
    Query Match:      18.68%     Indels:      29
    DB:               14         Gaps:       5
    US-09-921-992-50 (1-372) x CB619369 (1-853)
    Qy 74 LeuIleLysGlnValAsnValProLeuValAlaAspIleHisPheAspTyrArgIle 93
    Db 8 CTTGTTCAAGAAGATTACAACATCCCTAGTGGCTGATATTCATTTGCCCGACAGTT 67
    Qy 94 AlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsnProGlyAsnIleGly 113
    Db 68 GCTTTAAGAGTGGCTGAATGC---TTTGACAAAATTCGTGTCAACCCAGGGAATTTGCT 124
    Qy 114 AsnGlu----- 115
    Db 125 GATCCCGCTGCCAATTTGAGCAGCTTGAATATATCTGAAGATGATTATCAAAAAGAGCTT 184
    Qy 116 GluArgIleArgMetVal-----ValAspCysAlaArgAspLysAsnIlePro 131
    Db 185 GAGCATATCGAGAAGTCTTCTCCCGTTGGTTGAGAAATGCAAGCAGTATGGAAGACA 244
    Qy 132 IleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLysTyrGlyGlu 151
    Db 245 ATGCGGTATAGGAACAAATCATCGAAGTCTGCTGACCGCATATAGTATTACTATGTGTAT 304
    Qy 152 ProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeuAspArgLeu 171
    Db 305 ---TCTCCACGGGAATGCTTGAAGTCTGCTTTGGAAATTTGCCAGGATCTGTCCGAAGCTG 361
    Qy 172 AsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAlaValGluSer 191
    Db 362 GACTTCCATAACTTGTGTTTCAATGAAGCAAGTAACCCCTGTTATCATGTTGGTCCAAGCA 421
    Qy 192 TyrArgLeuLeuAlaLysGlnIle-----AspGlnProLeuHisLeuGly 206
    Db 422 TATCCGCTGCTGTAGCAGAAATGTATAACCTAGGGTGGGATTTATCTTTGCATCTGGGA 481
    Qy 207 IleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeu 226
    Db 482 GTTACAGAAGCTGGAGAGGGTGAAGATGGAGAGTGAAGTCTGCCATTGGCAACA 541
    Qy 227 LeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGlu 246
    Db 542 CTTCTGATGGATGGCTTGGCGGATACAAATCCGTGTCTCCCTCACGGAACCACTTGAAGAA 601

```

```

Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aaa cga cgg cca gta g
BACKWARD: gga aac agc tat gac cat g
Plate: 03 row: 0 column: 03
Seq primer: gta aaa cga cgg cca gta g.

```

FEATURES

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 23, 2003, 17:20:49 ; Search time 2600 Seconds
(without alignments)
3477.410 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866

Sequence: 1 MHNAPIQRKKSTRIYGVN.....RAKASQLDEARRIDVQVEK 372

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09921992/runat_21112003_162546_18418/app_query.fasta_1.519
-DB=EST -Qfmt=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=570 -MINLEN=0 -MAXLEN=2000000000
-USER=US09921992_@CGN_1_1_2810_@runat_21112003_162546_18418 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_htc.*
9: gb_est1.*
10: gb_est2.*
11: gb_htc.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pin.*
20: em_gss_vit.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_rod.*
26: em_gss_phg.*
27: em_gss_vrl.*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description	
1	820.5	44.0	1323	29	BZ575759	BZ575759 msh2_4614	
2	379	20.3	828	14	CB669733	CB669733 OSJNE02H	
3	348.5	18.7	853	14	CB619369	CB619369 OSIIEa030	
4	326.5	17.5	1442	11	AY104363	AY104363 Zea mays	
c	5	321.5	17.2	1268	29	BZ569685	BZ569685 pacs2-164
	6	315.5	16.9	579	12	B1920888	B1920888 EST540823
	7	313	16.8	720	28	BH235010	BH235010 PH_05.X.S
c	8	307.5	16.5	658	28	AZ570993	AZ570993 281PvD07
	9	276.5	14.8	852	14	CB628479	CB628479 OSIIEB04B
	10	271	14.5	602	14	CA022320	CA022320 H242M11r
11	269.5	14.4	610	12	BJ481080	BJ481080 BJ481080	
12	265	14.2	706	10	BG591263	BG591263 EST499105	
13	264	14.1	489	6	AU186794	AU186794 Porphyra	
14	262	14.0	601	9	AW042702	AW042702 ST3E04.P	
15	261	14.0	603	12	BJ551332	BJ551332 BJ551332	
16	260	13.9	558	9	AV917069	AV917069 AV917069	
17	260	13.9	608	12	BJ465443	BJ465443 BJ465443	
18	260	13.9	616	9	AV934187	AV934187 AV934187	
19	259	13.9	484	13	BQ080994	BQ080994 8an12Q09.	
20	259	13.9	651	9	AU252386	AU252386 AU252386	
21	257	13.8	640	13	BU091269	BU091269 6t70612.Y	
22	252	13.5	579	28	AZ523878	AZ523878 224PB502	
c	23	250.5	13.4	351	13	BU004160	BU004160 QGG37N12.
	24	244	13.1	465	9	AI482869	AI482869 EST242192
	25	236	12.6	576	13	BQ975044	BQ975044 QHI18D05.
26	232.5	12.5	545	9	AV626844	AV626844 AV626844	
27	229	12.3	532	10	BE924238	BE924238 EST428007	
28	229	12.3	628	13	BU048059	BU048059 PP_LEa003	
29	229	12.3	680	12	BG890681	BG890681 EST516532	
30	229	12.3	700	12	B1933447	B1933447 EST553336	
31	229	12.3	728	10	BG525868	BG525868 53-47-1-2	
32	228	12.2	423	13	BQ862800	BQ862800 QGC22B19.	
33	225	12.1	792	14	CB620208	CB620208 OSIIEa05D	
34	225	12.1	821	14	CB291771	CB291771 UCRS01.0	
35	225	12.1	821	14	CB628414	CB628414 OSIIEB03P	
36	225	12.1	1127	29	BZ577824	BZ577824 msh2_5589	
37	224	12.0	687	14	CA781720	CA781720 046F07AF	
38	223.5	12.0	413	12	BM059608	BM059608 KS01002E0	
39	222.5	11.9	544	13	BQ591926	BQ591926 E012583-0	
40	222.5	11.9	731	10	BG124857	BG124857 EST470503	
41	222	11.9	640	9	AW738440	AW738440 EST339867	
42	218.5	11.7	770	14	CB662235	CB662235 OSJNE06B	
43	216	11.6	448	9	AV644003	AV644003 AV644003	
44	216	11.6	482	9	AV642898	AV642898 AV642898	
45	216	11.6	502	9	AV626792	AV626792 AV626792	

ALIGNMENTS

RESULT	1	1323 bp	DNA	linear	GSS 17-DEC-2002
BZ575759	BZ575759	msh2_4614.x1 msh	Pseudomonas aeruginosa	genomic clone msh2_4614,	
LOCUS	BZ575759	genomic survey sequence.			
DEFINITION	BZ575759				
ACCESSION	BZ575759.1	GI:27210820			
VERSION	GSS.				
KEYWORDS	Pseudomonas aeruginosa				
SOURCE	Pseudomonas aeruginosa				
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;				
REFERENCE	1 (bases 1 to 1323)				

597	ATGAGAGCTTACAGGATGCTTCGCCCTCTGTGTGATCTATCCTTTCCTATTGGGGGTTACG	656
Qy	GluaLaGlyGlyAlaArgSerGlyAlaValIysSerAlaIleGlyLeuGlyLeuLeuLeu	228
Db	657 GAGCGGGGAATCTTTTAGCTCCAGTATCAAAATCCGCTATGCGTTTAGGGGGCTTTTA	716
Qy	229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle	248
Db	717 ATGAGAGGCGCATTTGGGATACGATCGCGGTATCCATCACAGGGGAATTAGAAAAATGAATC	776
Qy	249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle	268
Db	777 AAGTGCCACAGCAATTTTACGCCATAGCGGGCGGTGAAAGAGGGATTAATTGGATT	836
Qy	269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu	288
Db	837 TCITGCCCCACTCGCGGCGCATTTGAAGCCAATTTAGTGGATATGGCGATCAAGGTAGAA	896
Qy	289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn	308
Db	897 AAACGCTTAAGCCACATCAAAACCCCTTTAGACATTAGCGTGTAGTGGTTCGCTGGTGAAT	956
Qy	309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly	328
Db	957 GCTTTGGTGAAGCCAAGCATGCGATCGCTTTTGGGAATCGCAGCGGTTTG	1016
Qy	329 LeuTyrgluAspGlyValAlaGlyAspArgLeuAspAsnAsnAspMetIleAspGlnLeu	348
Db	1017 ATCATTAAGAGGGTAAAGTCATTCACAACTGGCTGAAAGGATTATTTTGAAACITTT	1076
Qy	349 GluaLaArgIleArgAlaLysAlaSerGlnLeuAspGluAla	362
Db	1077 GTCATGAAGTGGAAAAATTTAGCTTAAAGAAAGAGAAAAAAGT	1118

Search completed: November 23, 2003, 17:55:54
Job time : 2022 secs

```
QY 129 AsnileProileArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
DB 304 AATATTCCAATCGTATGGGTATCTGCTGGTAGTTAGAAAGAAATATTCAAAA 363
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 364 TACGGTTACCTACTGCTGAAGCAATGGTAGAAAGTGCATTCATCATATTAAATTC 423
QY 169 AspArgLeuAsnPhaAspGlnPhaLysValSerValLysAlaSerAspValPheLeuAla 188
DB 424 GAAGATTATAGATTTTATGATATCATTTGTTCTTTGAAAGCTTCTCATGTTAGCA 483
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnLeuAspGlnProLeuHisLeuGlyIleThr 208
DB 484 ATTGAAGCTTATGATTAAGCTAGTCGGCATTTAATTATCTCTACATCTTGGAAATTA 543
QY 209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
DB 544 GAATCTGGTACGCAGTTTCTGGAGGAATAAAAGTCTGCTGTTTAGGAGCGATCTC 603
QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
DB 604 AGTTTCGGCATTTGGAATACATTACGGGTATCTTTGAGTCTGATCCTCTGGAAGAGATA 663
QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
DB 664 AAAGTGGCCGGGAGGTTTAAATCGTTTGGCTTCTCTGATCCGCCATGCTTATC 723
QY 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAsnAlaLeuGlu 288
DB 724 TCCTGCGCTACTTTCGGTTCGAATAGAGATTGATTGATTCGATCCGTAATGAAGTGGA 783
QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
DB 784 AATTACATAGCAACGATTAAGCTCCGATTAAAGTAGCCGTGCTGGCTGTCGGGTCAAC 843
QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyCysAsnLysSerGly 328
DB 844 GTTCCGGAGAGCTCCGAGCAGATATCGGAATTGCTGGTTCAATGGAAGGCCCTT 903
QY 329 LeuTyrGluAspGly-----ValArgLys 336
DB 904 CTTTTTAGACATGGTAAATATCATCCGAAA 933
RESULT 15
AA14017
ID AA14017 standard; DNA; 1180 BP.
XX
AC AA14017;
XX
AC AA14017;
XX
DT 31-MAR-1999 (first entry)
XX
DE H. pylori GHPO 76 gene.
XX
KW GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease; ss.
XX
OS Helicobacter pylori.
XX
FH Key Location/Qualifiers
FT 51..1130
FT CDS /*tag = a
XX
XX
PN WO9843478-A1.
XX
PD 08-OCT-1998.
XX
PF 01-APR-1998; 98WO-US06371.
XX
PR 29-JUL-1997; 97US-0902615.
PR 01-APR-1997; 97US-0833457.
PR 24-JUN-1997; 97US-0881227.
```

```
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX (INMR ) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
XX WPI; 1998-542293/46.
DR P-PSDB; AAW98298.
XX
XX New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
XX
PS Claim 1; Page 372-374; 2054pp; English.
XX
CC This sequence represents a polynucleotide of the invention. It was
CC isolated from Helicobacter pylori and encodes a H.pylori GHPO protein.
CC The polypeptides can be used for preventing or treating Helicobacter
CC infections, and gastroduodenal diseases associated with these
CC infections, including acute, chronic, and atrophic gastritis, and peptic
CC ulcer diseases, e.g. gastric and duodenal ulcers. They can also be used
CC for the production of antibodies. The products can also be used for
CC detection and diagnosis.
XX
SQ Sequence 1180 BP; 380 A; 206 C; 289 G; 305 T; 0 other;
XX
Alignment Scores:
Pred. No.: 1.21e-72 Length: 1180
Score: 794.00 Matches: 162
Percent Similarity: 66.3% Conservative: 73
Best Local Similarity: 45.76% Mismatches: 117
Query Match: 42.55% Indels: 2
DB: 19 Gaps: 2
XX
US-09-921-992-50 (1-372) x AA14017 (1-1180)
QY 9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
DB 63 AGAGTTAAGACCAACCAATTTTATCGGTGGCGTAGGCCATAGGGGGTGATCTCCATA 122
QY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
DB 123 AGCACGCAAGCATGACCTTTAGCAAAACCGCTGATATTGAAAGCACTAAAAATCAAAT 182
QY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
DB 183 GACAGACTCAAACTCGCGGGCGCGATTTAGTAGAGGTGGCGGTAGTAATCAAAAGGAC 242
QY 69 AlaGluAlaPheLeuLeuLysGlnValAsnValProLeuValAlaAspIleHis 88
DB 243 GCTTAGCCTTAAAGAAATTGAAAGAAAGTGTCCCTTTTGCCTTTAATCGCTGATATTCA 302
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
DB 303 TTCATTATAATTCGCTCTC---ATTGCCGCTCAAGCGTGGATCGCATCAGATTAAAC 359
QY 109 ProGlyAsnIleGlyAsnGluArgIleArgMetValValAspCysAlaArgAspLys 128
DB 360 CCCGAAACATCGGCTCTAAAGAGAAGATCAAGCGGTGGTTGATGCTTTGTAAGAAAAA 419
QY 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
DB 420 AACATTCTTAAAGAAATTGGCGTGAATCTGGGAGTTTAGAAAAAGCAGTTTGTATCAAAA 479
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 480 TAGCGA---CCACCCCAAGGCGATGTAGAAAGCGCTTTGTATAACGCCCAACTTTA 536
QY 169 AspArgLeuAsnPhaAspGlnPhaLysValSerValLysAlaSerAspValPheLeuAla 188
DB 537 GAAGATTTGGATTTTACCAATTTTAAAGATTTCTTTAAAGCGAGCGATGATTCGCCACC 596
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
```


Db 118881 CTGCGCAGTTCAGCAGATGCGGAACTGACCGCGCGCTGCGACATCGTGGCGGTG 118940
Qy 62 SerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsnVal 81
Db 118941 GCGGTGCGGAGTCAGGACGATCCGACGACATCGCGGAAATTCGCGAAATCGCGGATT 119000
Qy 82 ProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGly 101
Db 119001 CCGGTTCATCGCGATATCCACTTCAGTCCCAATACGTTCAGGCCATCGACCGCGC 119060
Qy 102 ValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet--- 120
Db 119061 TCGCGGCGCGTGGCGTGAACCCGCGCAACATCCGTAAGTTCGACGAAGTCGCGCGGAC 119120
Qy 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
Db 119121 ATCTGCAAGCGCGCCACCGACGAGCATCTCGTGGTATCGGTGTGAACCGCGCTCG 119180
Qy 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
Db 119181 CTGGACAGGAATCTACCCCAAAATACGGTGGCCGACTCCGGAAGCGTGGTGGCTTC 119240
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
Db 119241 GCCTTGAAGGAAGCGCACATGTTGCGAGGAGCTGGCTTCCAGATTTCAGATCTCGTC 119300
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 119301 AAACACCATGAGTCTATCCATCCATGTTGCGAAACGATCCGACTGCTCTCCAAAGGCGAT 119360
Qy 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValLysSer 220
Db 119361 TGGCCCTCGACTCGCGCTGACCCGAGGCGGCGCTGCTGGCAAGCGCACCATCAATCC 119420
Qy 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
Db 119421 TGGCTGGCTGCTGGCTGCTTGGCGGAGGCGATTGGCGACAGCATTCGGGTGCTCTC 119480
Qy 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLysSerLeuArgIle 260
Db 119481 TCGCGCGCGCGCGCAAGAGGTCGAAAGTGGGTGCAAACTCTTGAATACATGGGATTG 119540
Qy 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
Db 119541 CGCCCCGCAAGTTCGACATCATCTCTGCGCGAGCTGCGCGCGCGCCCAAGTGGATGG 119600
Qy 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
Db 119601 ATCCAGTCCGCTCCGCGTCACCGAAGGCGCTGAAGGACGTGACCGCGCGATCCGAGTG 119660
Qy 301 SerIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
Db 119661 GCGCTATGGCTGCTGCTGCTCAACGGTTCGCGGCGAGCGACCGCAAGCGCGCTCGCGGTG 119720
Qy 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGly-----ValArg 335
Db 119721 GCTCCGCGCAATGGCAAGGCCAGATCTTCATCAAGCGCGAAGTTCATCAAGCGCGTGC 119780
Qy 336 LysAspArgLeuAspAsn-----AsnAspMetIleAspGlnLeuGlu 349
Db 119781 GAAGACCATGCTCGACACGCTGCTCACCATAGCGAAGCATATCGCGCGCCCAATGGAG 119840
Qy 350 Ala 350
Db 119841 GCC 119843
RESULT 14
ID ABQ68324 standard; DNA; 994 BP.
XX
AC ABQ68324;
XX
DT 29-AUG-2002 (first entry)

XX Listeria monocytogenes 4b contig DNA sequence #1090.
DE Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX Listeria monocytogenes 4b.
OS WO200228891-A2.
XX 11-APR-2002.
XX 04-OCT-2001; 2001WO-FR03061.
XX 04-OCT-2000; 2000FR-0012697.
XX (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX Kunst F, Glaser P;
XX WPI; 2002-332479/37.
XX New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators
XX Claim 14; SEQ ID 1137; 180pp; French.
XX The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ffp.wipo.int/pub/published_pct_sequences.
XX Sequence 994 BP; 296 A; 169 C; 239 G; 290 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 1 3e-74 Length: 994
Score: 812.00 Matches: 164
Percent Similarity: 70.32% Conservative: 54
Best Local Similarity: 52.90% Mismatches: 90
Query Match: 43.52% Indels: 2
DB: 24 Gaps: 1
US-09-921-992-50 (1-372) x ABQ68324 (1-994)
Qy 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
Db 4 GCTGGAAGAGCATGACTACTCAAGCTACATGATGTGCAAGCAACAGTAGCGGAATT 63
Qy 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
Db 64 CACCGATTAGAAGAGGTGGTGTGTCAGATTGTCGAGTTGCTTGTCTGATGAGCGTGCA 123
Qy 69 AlaGluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAspIleHis 88
Db 124 GCCAATGCTCTTAGTGCCATCAAGAAAGAGTTCATATTCCTCTTGTGGCAGATATTCAT 183
Qy 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
Db 184 TTTGATTACCGACTAGCAGCTTAAAGCTATTGATGACAGGTGTTGCAAAAATCCGGAATAT 243
Qy 109 ProGlyAsnIleGlyAsnGluAlaArgIleArgMetValValAspCysAlaArgAspLys 128
Db 244 CTGTGTAACATCGTCCGCGTGTGATCGGTGGAAGAGTGGTAAATGCTCTAAGCAAAA 303

QY 29 AlaValGlnSerMetThrAenThrArgThrThrAspValGluAlaThrValAenGlnIle 48
DB 84 GTCAATCAAGCATGACACAAACGAAACACATGAGCTTGACCAACCGTCGCGAATC 143
QY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
DB 144 AACAGACTCGCGAAGCAGGATGTCAAATCGTCGCGGTGCGCTGCTGATGAACGGGCT 203
QY 69 AlaGluAlaPheLysLeuIleLysGlnGlnValAenValProLeuValAlaAspIleHis 88
DB 204 GCGAGCGCATTCAGAGATCAAAAGCGGATATCCATCCTCTCTCGTGGATATTCAT 263
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLysArgIleAsn 108
DB 264 TTCAACTATAATGGCATTAAGCGATCGAAGCGGAGCGCCGATAAATCCGCATCAT 323
QY 109 ProGlyAsnIleGlyAenGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
DB 324 CCGGGTAACATCGCGCGCGGAAAGGTTGAAGCGTCTCAACGACGAGGAAGAAAG 383
QY 129 AsnIleProIleArgIleGlyValAenAlaGlySerLeuGluLysAspLeuGlnLys 148
DB 384 GGCATTCCGATCGGATCGCGTCATGAGGCTCTCGGAAACGAATCCTTGAGAG 443
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
DB 444 TACGGTATCCGACACGACGACGATGTCGAAAGCGCTCGACACATTAATTCCT 503
QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
DB 504 GAGGATCTCGATTTCACGATATCATCGTCAGCATGAGCGTCTGATGTAACCTGGCG 563
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyThr 208
DB 564 ATTGAGGCATATCAAAAGCGCTAAAGCTTTTCGATTATCCGCTTCATTAGGCATCAC 623
QY 209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
DB 624 GAATCGGAAACGCTGTTCCGGTACAGTGAAGGCGCGCGGTCTCGCGGATCCTT 683
QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
DB 684 TCAAAAGGAATCGGCAATCATTCGCGATTCTTTAAGCGCGACCGCTCGAAGAAGTA 743
QY 249 LysValGlyPheAspIleLysSerLeuArgIleArgSerArgGlyIleAsnPhelle 268
DB 744 AAAGTCGCGAGGAGCTCTGAAATCTTTTCGGGCTGCTTCAATGCGGCAACATTGATT 803
QY 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAlaAlaLeuGlu 288
DB 804 TCTGCGCGACCTTGGCGCGGATCGAATCGATTGATTCGATTCGCAATGAATTCGAA 863
QY 289 GlnArgLeuGluAspIleThrProMetAspValSerIleIleGlyCysValValAsn 308
DB 864 GATTACATCGGAAATCAAGCGCGCATCAAGTTGCGGTTCTCGGCTGTCGCTCAAC 923
QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLys 325
DB 924 GGTCCCGGAGAGCGCGGAGCGGATATCCGATGTCGCGCGCGCGCGCGG 974

RESULT 13

ABQ81842
ID ABQ81842 standard; DNA; 349980 BP.
XX
AC ABQ81842;
XX
DT 19-NOV-2002 (first entry)
XX
DE Bifidobacterium longum NCC2705 genomic sequence SEQ ID NO:1.
XX
KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
antidiarrheic; antibacterial; inhibitor of Salmonella; detection;

QY 3 AenGlnAlaProIleGln---ArgArgLysSerThrArgIleTyrValGlyAsnValPro 21
DB 118761 AGCGAATCGCCTTGCCCGCGCGCAAGTCCCGCGCATCATGCTGGTCCGGTGGCG 118820
QY 22 IleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAspVal 41
DB 118821 GTGGCGGGGAGCGCCCATCTCCGTGAGTGCATGATGATGATGATGATGATGATGAT 118860
QY 42 GluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgVal 61

Alignment Scores:
Pred. No.: 1,95e-71 Length: 349980
Score: 814.50 Matches: 167
Percent Similarity: 65.65% Conservative: 70
Best Local Similarity: 46.26% Mismatches: 111
Query Match: 43.65% Indels: 13
DB: 24 Gaps: 4

US-09-921-992-50 (1-372) x ABQ81842 (1-349980)

Sequence 349980 BP; 72540 A; 102738 C; 103221 G; 71481 T; 0 other;

The present invention describes a polynucleotide (I) comprising a sequence of a Bifidobacterium genome selected from the nucleotide sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at least 90% identity or which hybridises with the sequences given in ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding a fusion protein, comprising a sequence selected from 1097 sequences given in ABP65258 to ABP65354 ligated in frame to a polynucleotide encoding a heterologous polypeptide. (I) has antidiarrheic and antibacterial activities, and can be used as an inhibitor of Salmonella. (II) (which is a probe) is useful for the detection and/or identification of Bifidobacterium longum in a biological sample. A carrier containing the lactic acid bacterium Bifidobacterium longum NCC2705 (CNCM I-2618) can be used for preventing and/or treating diarrhoea brought about by pathogenic bacteria and/or rotavirus. The carrier is a food composition selected from milk, yogurt, curd, cheese, fermented milks, milk based fermented products, ice-creams, fermented cereal based products, milk based powders, infant formula, pet food or a pharmaceutical composition selected from tablets, liquid bacterial suspensions, dried oral supplement, wet oral supplement, dry tube feeding or wet tube feeding. (I) is useful in DNA arrays or chips to carry out analysis of the expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent Bifidobacterium related nucleotide sequences given in the Sequence Listing from the present invention but not mentioned further within the specification.

N.B. The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied by the European Patent Office.

```
QY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
Db 3183 ACTATCCAAAGCATGACTACTACAAGACACATGATGTCGAGCACAGTACGCGAAAT 3242
QY 49 LysAlaLeuGluArgValGlyAlaSerIleValArgValSerValProThrMetAspAla 68
Db 3243 CACCGATTAGAAGAGGTGGTGTGCAGATTGTGCGAGTTGCTGTCTGATGAGCGTGCA 3302
QY 69 AlaGluAlaPheLysLeuIleLysGlnGlnValAsnValProLeuValAlaSerIleHis 88
Db 3303 GCCAATGCTCTTAGTCCCATCAAGAAAGGATTCAATTCCTTCTGTGGCAGATATTCAT 3362
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
Db 3363 TTTGATTACCGACTAGCACTTAAAGCTATTGATGCAGGTGTTGCACAAATCCGGATTAT 3422
QY 109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysAlaArgAspLys 128
Db 3423 CCTGGTAAATCGGTCCGCTGATCGGTGGAAGAAAGTGTAAATGCTGCTAAAGCAAAA 3482
QY 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnLys 148
Db 3483 AATATTCCAAATCCGATTGGGTGTTAATGCTGGTAGTTTGAAGAAATATTATCAAAA 3542
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
Db 3543 TACGGTTACCTTACTGCTGAGGAATGGTAGAAGTGCATTGATCATATTAATAATCTC 3602
QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db 3603 GAAGATTAGATTATTTATGATATCATTTGTTCTTTGAAAGCTTCTGATGTGAATTAGCA 3662
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db 3663 ATTGAAGCTTATGATAAAGCTAGTCGCGCATTTAATTATTCCTCATCTTGAATATCA 3722
QY 209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
Db 3723 GAATCTGGTACCGAGTTGCTGGAGCAATAAAGTGTCTGCTGTTTGGAGCGATACTC 3782
QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
Db 3783 AGTTTGGCATTTGGAATACATTTACCGGTATCTTTGAGTGCTGATCCTGTGGAAGAGATA 3842
QY 249 LysValGlyPheAspIleLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
Db 3843 AAAGTGGCCGGAGGTTTAAATCGTTTGGCCTTTCTCGAATGCCGCCATGCTTATC 3902
QY 269 AlaCysProThrCysSerArgGlnGluPheAspValIleGlyThrValAlaAlaLeuGlu 288
Db 3903 TCCTGCTTACTTTCGCTCGATAGAGATTGATTGATTCGATCCGTAATGAAGTGGA 3962
QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
Db 3963 AATTACATAGCAACGATTAAAGTCCGATTAAAGTAGCCGCTGCTGCTGCTGCGTCAAC 4022
QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysLysSerGly 328
Db 4023 GGTCGCGGAGAGCTCGGAGCAGATATCGGAATTTGCTGTTCAATGAGAGGCCCTT 4082
QY 329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeu 348
Db 4083 CTTTTTAGACATGTTAAATCATCCGAAAGTACCAGACCATTTATGTTAGAGAACTT 4142
QY 349 GluAlaArgIleArgAlaLysAlaSerGlnLeu 359
Db 4143 AAGAAAGGAATTGATATTTTGGCCAGAGAATA 4175
```

RESULT 12

ABK74401

ID ABK74401 standard; DNA; 1083 BP.

XX

AC ABK74401;

```
XX 13-AUG-2002 (first entry)
DT XX
DE DE Bacillus licheniformis genomic sequence tag (GST) #1692.
KW XX Differential gene expression; genomic sequenced tag; GST;
KW XX altered culture condition; environmental stress;
KW XX physiological provocation; ds.
XX XX
OS XX Bacillus licheniformis.
XX XX
PN WO200229113-A2.
XX XX
PD 11-APR-2002.
XX XX
PF 05-OCT-2001; 2001WO-US31437.
XX XX
PR 06-OCT-2000; 2000US-0680598.
PR 27-MAR-2001; 2001US-279526P.
XX XX
PA (NOVO ) NOVOZYMES BIOTECH INC.
PA (NOVO ) NOVOZYMES AS.
XX XX
PI Berka R, Clausen IG;
XX XX
DR WPI; 2002-416684/44.
XX XX
PT Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array
XX XX
PS Claim 4; SEQ ID NO 1692; 200pp; English.
XX XX
CC The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive
CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at
CC ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 1083 BP; 314 A; 278 C; 265 G; 226 T; 0 other;

Alignment Scores:
Pred. No.: 1.34e-75 Length: 1083
Score: 822.00 Matches: 162
Percent Similarity: 68.45% Conservatives: 55
Best Local Similarity: 51.10% Mismatches: 100
Query Match: 44.05% Indels: 0
DB: 24 Gaps: 0

US-09-921-992-50 (1-372) x ABK74401 (1-1083)
QY 9 ArgArgLysSerThrArgIleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
Db 24 CGTACAAAACGCGTCCCGCTTAAAGTGGGACCTTTAAACAATAGCGGCAATACGAAGTC 83
```

```
Query Match: 46.52% Indels: 2
DB: 24 Gaps: 2
US-09-921-992-50 (1-372) x ABA03041 (1-2944528)

QY 10 ArgLysSerThrArg---lleTyrValGlyAsnValProIleGlyAspGlyAlaProfile 28
Db 1474820 CGCGAAACACATCGCCAGTCCCAAGTGGGTAAATTTAACTATTGGTGTAGTGGAGTAATTA 1474879

QY 29 AlaValGlnSerMetThrArgThrArgThrArgValGluAlaThrValAsnGlnIle 48
Db 1474880 ACTATCCAAGCATGACTACTCAAAAGACACATGATGTGCAAGACACATGACAGAAATTT 1474939

QY 49 LysAlaLeuGluArgValGlyAlaAspIleValArgValSerValProThrMetAspAla 68
Db 1474940 CACCGATTAGACAGACTGGTGTCTAGATTGGGAGTGTCTGTCTCTGATGACCTGCA 1474999

QY 69 AlaGluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAspIleHis 88
Db 1475000 GCGAATGCTCTTAGTCCCATCAAGAAAAAGATTTCATATTCGCTTGTGCGACATATTCAT 1475059

QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
Db 1475060 TTGTGATTACCGACTAGCACTTAAAGCTATTGATGCGAGGTGTTGACAAAAATTCGATTAT 1475119

QY 109 ProGlyAsnIleGlyAsnGluAluArgIleAlaGluMetValValAspCysAlaArgAspLys 128
Db 1475120 CTGTGTAACTATGCTCGCCGATCGGTGGGAAAGTGGTAAATGCTGCTAAAGCAAAA 1475179

QY 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
Db 1475180 AATATTCCAATCCGTATGCGGTAAATGCTGTAGTTAGTAAAGAAAAATTTATCAAAA 1475239

QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
Db 1475240 TATGGTTACCTCTACTGCTACCGAATGGTAGAAAGTGCACTTCCCATATTTAAATTTCTC 1475299

QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db 1475300 GAAGATTAGATTTTATCATATCATCTTTTCTTTGAAGCTTCTGATGTGAATTTAGCA 1475359

QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db 1475360 ATTGAAGCTTATGATAAGCTAGTCGCGCATTTAATTTATCTCTGATCTCGGAATTACA 1475419

QY 209 GluAlaGlyGlyAlaArgSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeu 228
Db 1475420 GAATCTGGTACACAAATTTCTCGAGGAATAAAAGTGTCTGTGTTTAGGAGCGGATCTC 1475479

QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAspProValGluGluIle 248
Db 1475480 AGTTTGGGCAATTTGGAATAATACATACAGATATCTTTGAGTGCTGATCTCTGTGGAAGAAATA 1475539

QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
Db 1475540 AAAGTGGCCGCGGAAGTTTAAATCATTTGCTCTTCTCGAAATGCCCAATGCTTATC 1475599

QY 269 AlaCysProThrCysSerArgGlnGlnPheAspValIleGlyThrValAsnAlaLeuGlu 288
Db 1475600 TCCTGCTCTACTTGGGTGCGATAGATGATTGATTAAATTCGTATCGCTAATGAAGTGGAA 1475659

QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
Db 1475660 AATTACATAGCAAAAGATTGAAGTTCGATTAAGTAGCGCTGCTGTGCTGCGGTCAAC 1475719

QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyAsnLysLysSerGly 328
Db 1475720 GGCCCTGGAGAGCTCGCAAGCCGATATCGGAATTTGCTGTTCAACCGGAAAGCGCTT 1475779

QY 329 LeuTyrGluAspGlyValArgLysAspArgLeuAspAsnAsnAspMetIleAspGlnLeu 348
Db 1475780 CTTTTTAGACATGGTAAAAATTTATCGAAAGATCCGGAAGCTATTATGATGACCAACTT 1475839
```

```
QY 349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAspValGln 368
Db 1475840 AAGAAAGAAATTTGATATTTTTCGACAGAATTT---TTTGTGAAGAAATAGATTTGGAA 1475896

QY 369 GlnVal 370
Db 1475897 AGCCTT 1475902

RESULT 11
ABQ70939
ID ABQ70939 standard; DNA; 6157 BP.
XX ABQ70939;
XX 29-AUG-2002 (first entry)
DE Listeria monocytogenes 4b contig DNA sequence #881.
XX Antibacterial; Listeria; food contamination; mutational analysis;
KW infection; ds.
XX Listeria monocytogenes 4b.
XX WO200228891-A2.
XX 11-APR-2002.
XX 04-OCT-2001; 2001WO-FR03061.
XX 04-OCT-2000; 2000FR-0012697.
XX (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX Kunst F, Glaser P;
XX WPI; 2002-332479/37.
XX New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
Claim 14; SEQ ID 3752; 180pp; French.
XX The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at fcp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6157 BP; 1731 A; 1065 C; 1209 G; 2150 T; 2 other;
```

```
Alignment Scores:
Pred. No.: 1,12e-78 Length: 6157
Score: 861.50 Matches: 175
Percent Similarity: 68.66% Conservative: 66
Best Local Similarity: 49.86% Mismatches: 109
Query Match: 46.17% Indels: 1
DB: 24 Gaps: 1
```

US-09-921-992-50 (1-372) x ABQ70939 (1-6157)

```
QY 10 ArgLysSerThrArg---lleTyrValGlyAsnValProIleGlyAspGlyAlaProfile 28
Db 3123 CGCGAAACACATCGCCAGTCCCAAGTAGGTAATTTAACTATTGGTGTAGTGGAGTAATTA 3182
```



```
SQ Sequence 1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;

Alignment Scores:
Pred. No.: 2,34e-80 Length: 1107
Score: 868.00 Matches: 178
Percent Similarity: 67.68% Conservative: 67
Best Local Similarity: 49.17% Mismatches: 115
Query Match: 46.52% Indels: 2
DB: 24 Gaps: 2

US-09-921-992-50 (1-372) x ABQ67966 (1-1107)

QY 10 ArgLysSerThrArg---lleTyrValGlyAsnValProIleGlyAspGlyAlaProIle 28
Db 19 CGCGAAACACCTCGCCCGCTCCCAAGTGGCTAATTAACTATTGGTGTAGTGAGGAATTA 78
QY 29 AlaValGlnSerMetThrAsnThrArgThrThrAspValGluAlaThrValAsnGlnIle 48
Db 79 ACTATCCAAAGCATGCTACTACTCAAGACACATGATGTCGAAGCAACAGTAGCAGCAAAAT 138
QY 49 LysAlaLeuGluArgValGlyAlaAspIleValAtgValSerValProThrMetAspAla 68
Db 139 CACCGATTAGAGAAGCTGGTGTGAGATGCGGAGTTGCTTGCTCCTGATGAACGTGCA 198
QY 69 AlaGluAlaPheLysLeuIleLysGlnValAsnValProLeuValAlaAspIleHis 88
Db 199 GCGAATGCTCTTAGTGCCATCAAGAAAAAGATCATATTCCGCTTGCGCAGATATTAT 258
QY 89 PheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAspCysLeuArgIleAsn 108
Db 259 TTGTGATTACCGACTAGCACTTAAAGCTATTGATGCAGGTTGTGACAAAAATTCGATTAT 318
QY 109 ProGlyAsnIleGlyAsnGluGluArgIleArgMetValValAspCysLeuArgAspLys 128
Db 319 CTGTGTAACTTTGGTCCGCTGCGGTGAGGAAAAAGTGGTTAATGCTGCTAAAGCAAAA 378
QY 129 AsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLysAspLeuGlnGluLys 148
Db 379 AATATTCCAATCGTATTGGGTTAATGCTGTGTAGTTAGAAAAGAAAATATTCAAAAA 438
QY 149 TyrGlyGluProThrProGlnAlaLeuLeuGluSerAlaMetArgHisValAspHisLeu 168
Db 439 TATGTTACTCCCTACTGCTGACGGAATGTGAGAAAGTGCACCTTGCCCATATTAAAAATCTC 498
QY 169 AspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSerAspValPheLeuAla 188
Db 499 GAAGATTATGATTTATGATATCATTTGTTCTTTTGAAGGCTTCTGATGTGAATTAGCA 558
QY 189 ValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeuHisLeuGlyIleThr 208
Db 559 ATTGAAGCTTATGATAAAGCTAGTCGCGCATTAATTATCTCTCGCATCTCGGAATTACA 618
QY 209 GluAlaGlyGlyAlaAspSerGlyAlaValLysSerAlaIleGlyLeuGlyLeuLeuLeu 228
Db 619 GAATCTGTGTACAAATTTGCTCGAGGAATAAAAGTGTGCTGGTTTAGGACGGATACTC 678
QY 229 SerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAspProValGluGluLeu 248
Db 679 AGTTGGGCATTTGGAATACATTACAGTATCTTTGAGTGCTGATCTGCTGGAAGAAATA 738
QY 249 LysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArgGlyIleAsnPheIle 268
Db 739 AAAGTGGCCCGGAAGTTTAAATCATTTGCTCTTCTCGAATCGCGCAATGCTTATC 798
QY 269 AlaCysProThrCysSerArgGlnGlnPheAspValIleGlyThrValAsnAlaLeuGlu 288
Db 799 TCCTGCCCTACTTGGGTGCAATAGATATTGATTAAATTCGTATCGCTAATGAAGTGAA 858
QY 289 GlnArgLeuGluAspIleIleThrProMetAspValSerIleIleGlyCysValValAsn 308
Db 859 AATTACATGACCAAGATTGAAGTTCGATTAAGTAGCCGTGCTGGTGTGCGGTCAAC 918
QY 309 GlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGlyAsnLysSerGly 328

1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;

Db 919 GGCCTCGAGAAGCTCGGAAGCCGATATCGGAATTGCTGTTCAACGGAAGGCCTT 978
QY 329 LeuTyrGluAspGlyValArgLysAspArgLysAspAsnAspMetIleAspGlnLeu 348
Db 979 CTTTTAGACATGTTAAATATTTCGAAAAGTACCGGAAGCTATTATGATACGAACTT 1038
QY 349 GluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAlaArgArgIleAspValGln 368
Db 1039 AAGAAAGAAATTTGATATTTTGGCAGAAATTT---TTTGTGAAGAAATAGATTGGAA 1095

369 GlnVal 370
1096 AGCCTT 1101

RESULT 9
ABQ69993
ID ABQ69993 standard; DNA; 1107 BP.
XX ABQ69993;
AC ABQ69993;
DT 29-AUG-2002 (first entry)
XX
DE Listeria monocytogenes EGDe DNA sequence #205.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
infection; ds.
XX
OS Listeria monocytogenes EGDe.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP ) INST PASTEUR.
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
WPI; 2002-332479/37.
XX
New genomic sequences from Listeria species, useful for detection,
treatment and prevention of infection, also related polypeptides,
antibodies and modulators
Claim 16; SEQ ID 2806; 180pp; French.
The present invention relates to nucleic acid sequences
(ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
and primers for identification and/or detection of Listeria (e.g. as
contaminants in foods, or mutational analysis) and for analysis of
gene expression. Proteins encoded by the nucleic acid sequences can be
used to screen for compounds that modulate gene expression, replication
and pathogenicity of Listeria (potential therapeutic agents), also for
treating infections by Listeria, and are useful as immunogens in
anti-Listeria vaccines.
Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic format
directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;

Alignment Scores:
Pred. No.: 2,34e-80 Length: 1107
Score: 868.00 Matches: 178
Percent Similarity: 67.68% Conservative: 67
Best Local Similarity: 49.17% Mismatches: 115
Query Match: 46.52% Indels: 2
DB: 24 Gaps: 2
```

CC AAF28514-AAF28554). The library has a number of uses described in the
 CC specification e.g. is useful for identifying diagnostic and therapeutic
 CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
 CC aerobic, gram-negative diplococcus, normally found among the bacterial
 CC flora of human upper airways. M. catarrhalis is known to cause acute,
 CC localised infections such as otitis media, sinusitis and bronchopulmonary
 CC infection and life-threatening, systemic diseases including endocarditis
 CC and meningitis.

XX SQ Sequence 33140 BP; 9472 A; 6496 C; 7384 G; 9787 T; 1 other;

Alignment Scores:

Pred. No.: 2,81e-109 Length: 33140
 Score: 1166.00 Matches: 220
 Percent Similarity: 78.92% Conservativeness: 72
 Best Local Similarity: 59.46% Mismatches: 76
 Query Match: 62.49% Indels: 2
 DB: 22 Gaps: 1

US-09-921-992-50 (1-372) x AAF28536 (1-33140)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
 DB 21741 ATACAAAT-----CGATTAAACCGCCCTTACCACCAAAAAATCATGTGGCAATGTC 21794
 QY 21 ProfileGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
 DB 21795 GCCATCGCGCGCATGCCATCCATCGGTGCAAGTATGACAAATACAAACACCTGTGAT 21854
 QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
 DB 21855 ATTGATCAACACGTTCGACAAATTCAGCGATGCGTGGACGACGATCGGACTGTGATCGT 21914
 QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnInValAsn 80
 DB 21915 GTATCGACACCAACGATGGAGCGTGGCTTGTGAGATAAAAAACCGGTATCC 21974
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
 DB 21975 ATTCCACTCATTCGCCATATTCATTGTGATCACAAATGCCATTCGATGGCGGATGTC 22034
 QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgMet 120
 DB 22035 GGTGACAGACTCGCTGATCAATCAATCAGGTATATTTGGCAATGATCAAAAAGTCAAAAGAA 22094
 QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
 DB 22095 GTGGTTGCACGCGCAGCCATCATATATGTCGCCGATTCGTATCGGTGGAATGCAGGCTCA 22154
 QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyProThrProGlnAlaLeuLeuGluSer 160
 DB 22155 CTTGAAAAGGATTTACAAAAAATAATCGAGACCGACAGGTGAGCGGATGCTAGATCG 22214
 QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
 DB 22215 GCGTTACGCGCACATGATATATTTAGAAAATCTTAATTTTCAGAAATATAAAATCTCAGTC 22274
 QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
 DB 22275 AAAGCTCCAAATGATTTTAAACCTTGATGCGGTATCGATTGATTTACGACCAAAATGAT 22334
 QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
 DB 22335 AATCCATTCGATCTTGGCTTACCGAAGCGGGGTATATCGCATGGCAGCGTCAATCT 22394
 QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
 DB 22395 GCCATTGCTTTGGCGGCTTTGCTACTTGTGATGCGATTCGCGATACCATCTCTCTTTC 22454
 QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLysSerLeuArgIle 260
 DB 22455 GCGGCGACGACCAAGAAAGAGATTAATAATTTGGCTTTTGACATTTCTAAATCACTGGGTATT 22514

QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
 DB 22515 CGTAGTAATGGCGTCAATTTTATTCTTGTCCAGTTGTAGCCGTCAAGAGTTTGATGTG 22574
 QY 281 IleGlyThrValAsnAlaLeuGlnArgLeuGluAspIleIleThrProMetAspVal 300
 DB 22575 ATCAAGTGTGAATGAGCTTGAAGCTCGTTAGAGGATATTCTGTGAGCCACTAGATCTG 22634
 QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
 DB 22635 TCAGTTATTTGGCTGTAAAGTCAATGGCCAGGTGAAGCCAAAGACGACGATCGGTGTC 22694
 QY 321 ThrGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
 DB 22695 GTGGGAGCCGCCCAATAGTTTGGTTTATTAACATGCTGCAAAAGCCATCTGATTGAT 22754
 QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
 DB 22755 ACCAAAAATTTGGTAGATGAGATTGAGCGGTGGTGGCTCAAGTCAAGACATTCAA 22814
 QY 361 GluAlaArgArgIleAspValGlnGlnVal 370
 DB 22815 GATAACGGCGCAATGAGATTATCCGTATA 22844
 RESULT 8
 ABQ67966
 ID ABQ67966 standard; DNA; 1107 BP.
 XX AC ABQ67966;
 XX DT 29-AUG-2002 (first entry)
 XX DE Listeria monocytogenes EGD DNA sequence #90.
 XX KW Antibacterial; Listeria; food contamination; mutational analysis;
 XX KW infection; ds.
 XX OS Listeria monocytogenes EGD.
 XX PN WO200228891-A2.
 XX PD 11-APR-2002.
 XX PF 04-OCT-2001; 2001WO-FR03061.
 XX PR 04-OCT-2000; 2000FR-0012697.
 XX PA (INSP) INST PASTEUR.
 XX PA (CNRS) CNRS CENT NAT RECH SCI.
 XX PI Kunst F, Glaser P;
 XX WPI; 2002-332479/37.
 XX DR
 XX PT New genomic sequences from Listeria species, useful for detection,
 XX PT treatment and prevention of infection, also related polypeptides,
 XX PT antibodies and modulators
 XX PS Claim 16; SEQ ID 779; 180pp; French.
 XX CC The present invention relates to nucleic acid sequences
 XX CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
 XX CC and primers for identification and/or detection of Listeria (e.g. as
 XX CC contaminants in foods, or mutational analysis) and for analysis of
 XX CC gene expression. Proteins encoded by the nucleic acid sequences can be
 XX CC used to screen for compounds that modulate gene expression, replication
 XX CC and pathogenicity of Listeria (potential therapeutic agents), also for
 XX CC treating infections by Listeria, and are useful as immunogens in
 XX CC anti-Listeria vaccines.
 XX CC Note: The sequence data for this patent did not form part
 XX CC of the printed specification, but was obtained in electronic format
 XX CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Claim 1; Page 16-230; 237pp; Japanese.

The present invention describes a gene (I) derived from *Buchnera* sp. containing the DNA (a) or (b), (a) has a fully defined base pair sequence selected from a table of sequences found in the *Buchnera* sp. genomic DNA of ABA92787 given in the specification or is a DNA selected from complementary DNA sequences, and (b) is a DNA which hybridizes with the DNA (a) and encodes a protein. Also described are: (1) a recombinant vector (II) containing (i); (2) a transformant (III) containing (II); (3) a genomic DNA of *Buchnera* sp. containing the sequence given in ABA92787; (4) a plasmid derived from *Buchnera* sp. containing DNA (c) or (d), (c) is a DNA containing a fully defined sequence given in ABA92788 or ABA92789 and (d) is a plasmid which hybridizes with a DNA; and (5) a method for the preparation of a protein in which (III) is cultured and the expression protein of the objective protein is collected from the resultant culture. The DNA is useful for developing agricultural chemicals for exterminating cockroaches. The present sequence represents the specifically claimed *Buchnera* sp. genomic DNA sequence, from the present invention.

Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;

Alignment Scores:

Pred. No.:	1-41e-119	Length:	640681
Score:	1282.00	Matches:	242
Percent Similarity:	83.88%	Conservative:	65
Best Local Similarity:	66.12%	Mismatches:	57
Query Match:	68.70%	Indels:	2
DB:	24	Gaps:	2

US-09-921-992-50 (1-372) x ABA92787 (1-640681)

QY	1	MethHisAsnGlnAlaPro---IleGlnArgAlaGlySerThrArgIleTyrValGlyAsn 19
DB	314272	ATGAATAATAAGTGTAAATATTCAATAGAGAAATCTGATCGTATTATGTGGAAAA 314331
QY	20	ValProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThr 39
DB	314332	GTTCGCTATTGGCAATTAATGGCCAAATATCAGTTCAATCTTGACAAATCTCGTACTACT 314391
QY	40	AspValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleVal 59
DB	314392	AATATCTCTGAACTATTAACTCAATCTTAGAGTTACAAAAAGTAGGATGATATTGTT 314451
QY	60	ArgValSerValProThrMetAspAlaGluAlaPheLysLeuIleLysGlnGlnVal 79
DB	314452	CGTATTCTTATACCCCAATTTAAAGCTGCAGATCAATTCAGAAATAAAAAACAAACA 314511
QY	80	AsnValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGlu 99
DB	314512	AATGTTCCATTGATTCAGATATACATTTTGATTACAGATTAGCTTTACAGCTATAAAA 314571
QY	100	TyrGlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArg 119
DB	314572	TATGTTGCAGATTGTTTAAAGATTAATCTCGGAATATTGGAAATAAAAGAGAGATCA 314631
QY	120	MetValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGly 139
DB	314632	GAAATCATTTCTTACCAAAAGATGAAATATTCCAAATTCGATTGCTGTTTAAATGCTGA 314691
QY	140	SerLeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGlu 159
DB	314692	TCTTTGAAAAGAGATATATATAAAAAATATAAATACCTTACTCCAGATGCATTAGTAGAA 314751
QY	160	SerAlaMetArgHisValAspHisLeuAspArgLeuAsnPheAsnGlnPheLysValSer 179
DB	314752	TCAGCTATGAGGCATATTCAATCTTGTGCTTTAAATTTTAAATCAATTTTAAAGTTAGT 314811
QY	180	ValLysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIle 199
DB	314812	GTTAAAGCTGCTGATGATTTTTTAGCTATTGATCATATTCGATGTTAGGAAAAAAATTT 314871
QY	200	AspGlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLys 219

DB	314872	ACACAACCTTTGCATATTGGCATAACTGAATCCGGTGGTTTAAGGAATGGAACAGTTAAA 314931
QY	220	SerAlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSer 239
DB	314932	TCATCTATAGGTATTGCTTTATTATTATTAGGAAGCATTGGAGATACAATACGAGTTTCA 314991
QY	240	LeuAlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArg 259
DB	314992	TTAGCGGCACATCCCACTGGAAGAGTAAAGTAGGTATGACATTTTAAAGTTTATCT 315051
QY	260	IleArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAsp 279
DB	315052	TTAAGAGCAAGAGGTATTAAATTTATTGCTTCTGCTTCTTCTAGACAAGAAATTTGAT 315111
QY	280	ValIleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAsp 299
DB	315112	GTAATTAATACAGTAAATCACTAGAAAAAATCTAGAAAGATATCTCGACTCCCATAGAT 315171
QY	300	ValSerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGly 319
DB	315172	GTATCAATTTATGTTGCTTGTGTTTATGGAATAGGTGAATCTTAAATAGCACTTTAGGT 315231
QY	320	ValThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArg 338
DB	315232	CTAGCAGGAAGTCATAAGAAAAAGTGCATTTTATGAAGACGGAGTAAGACAAAAAGAAAA 315291
QY	339	LeuAspAsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGln 358
DB	315292	ATAAAAACGAAGAAATTTATAGAAAAAATGGAATTTAAATTCGAAAAAATAAGACAAA 315351
QY	359	LeuAspGluAlaArgArg 364
DB	315352	TTAAATATTCAAAAAA 315369
RESULT 7		
ID	AAF28536	AAF28536 standard; DNA; 33140 BP.
XX	AAF28536;	
AC	AAF28536;	
DT	04-APR-2001	(first entry)
DE	Genomic fragment #23.	
XX	Genomic library; bacteria; human upper airway; otitis media; sinusitis;	
KW	bronchopulmonary; endocarditis; meningitis; ss.	
OS	Moraxella catarrhalis.	
XX	WO200078968-A2.	
PN	28-DEC-2000.	
XX	16-JUN-2000; 2000WO-US16649.	
PF	18-JUN-1999; 99US-0140121.	
XX	(INCY-) INCYTE GENOMICS INC.	
PA	Lagace RE, Patterson C, Berg KL;	
XX	WPI; 2001-041427/05.	
DR	Genomic library for identifying diagnostic and therapeutic	
XX	compositions, and for identifying virulence factors, regulatory	
PT	elements and drug targets, comprises Moraxella catarrhalis nucleic	
PT	acids -	
XX	Claim 1; Page 191-199; 545pp; English.	
PS	The present invention relates to a Moraxella catarrhalis genomic library	
CC	comprising of a combination of 41 nucleic acid molecules (see	

PR 07-JUN-1995; 95US-0487429.
 PR 21-APR-1995; 95US-0426787.
 PR 07-JUN-1995; 95US-0476102.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (UJO) UNIV JOHNS HOPKINS.
 XX
 PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
 XX WPI; 1996-485782/48.
 DR
 XX
 XX Haemophilus influenzae Rd genome recorded on computer readable
 PT medium - useful for identifying commercially important nucleic acid
 PT fragments by homology searching
 XX
 PS Claim 1; Page 77.2-77.1091; 1291pp; English.
 XX
 CC This sequence represents the complete genome sequence of the bacterium
 CC Haemophilus influenzae strain Rd. The invention relates to a computer
 CC readable medium (CRM) having recorded upon it the complete H. influenzae
 CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide
 CC sequence at least 99% identical to (I). By providing the full-length
 CC genomic sequence in a computer readable form, it is possible to identify
 CC commercially important nucleic acid fragments and expression modulating
 CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
 CC regulate the expression of a nucleic acid molecule. Vectors and altered
 CC organisms comprising the predicted ORFs can be used to produce any of the
 CC polypeptide fragments of the H. influenzae Rd genome.
 XX
 SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Alignment Scores:
 Pred. No.: 2,6e-152 Length: 1830121
 Score: 1603,50 Matches: 320
 Percent Similarity: 93,66% Conservative: 20
 Best Local Similarity: 88,15% Mismatches: 22
 Query Match: 85,93% Indels: 1
 DB: 17 Gaps: 1

US-09-921-992-50 (1-372) x AAT42063 (1-1830121)

QY 4 GlnAlaProIleGlnArgLysSerThrArgIleTyrValGlyAsnValProIleGly 23
 DB 390972 CAGCCCAACTATTAAAGCGTCGTGAATCGACAAAAATTTATGTGGAAATGTACCAATTTGGT 391031
 QY 24 AspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAspValGluAla 43
 DB 391032 GGGGATGCGCCTATTGCGGTGCAATCAATGACAAATCTCGCACCACTGATGTGAAGCG 391091
 QY 44 ThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArgValSerVal 63
 DB 391092 ACAGTTCCTCAATTAATCAATTAGAACGTTGTTGGTCAGATATTGTTGCTGTATCTGTT 391151
 QY 64 ProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsnValProLeu 83
 DB 391152 CCACCAATGGATGTCGGGAAGCATTTAAACAAATTAACAAAGTGAATGTTCCGCTC 391211
 QY 84 ValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyrGlyValAsp 103
 DB 391212 GTAGCAGATATTCAATTCGACTATCGTATCGCTTAAAGTCGAGATATGAGTGGAT 391271
 QY 104 CysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgMetValValAsp 123
 DB 391272 TGTTTACGTATCAATCTCGCAACATTTGTCGTGAAGATCGCGTCGCTGTTGAT 391331
 QY 124 CysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySerLeuGluLys 143
 DB 391332 TGTGCGCGACACAAAAATATTCGATTTCGATTGTTGGTGAATGCAGCTCTTTAGAAAAA 391391
 QY 144 AspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuGluSerAlaMetArg 163
 DB 391392 GATTTGCAAGAAAAATATGCGCAACCAACCGCAGACCTTTGTTAGAAATCCGATTTGCGT 391451

QY 164 HisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerValLysAlaSer 183
 DB 391452 CATGTAGAAATTTCTAGATCGTCTTAATTCGATCAGTTTAAAGTCAGCGTAAAGCGCTCC 391511
 QY 184 AspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAspGlnProLeu 203
 DB 391512 GATGTATTCTTAGCGGTTGAATCTTATCGTTTACTGGCTAAAGCAATTAACAGCGCTTTA 391571
 QY 204 HisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyValValLysSerAlaIleGly 223
 DB 391572 CATTTAGGCATTACAGAACGAGTGGCGCACCGCTGGTGCAGTAAATCTCGCAGTGGT 391631
 QY 224 LeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeuAlaAlaAsp 243
 DB 391632 TTAGGAATGTTATTAGCTGAGGCGATTGGCGATACACTACGCGTCTCTTTGGCGGAGAT 391691
 QY 244 ProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIleArgSerArg 263
 DB 391692 CCTGTAGAGGAATCAAAAGTCGGTTTGATATTTTGAATCTTTACGGATTTCGTTCAAGA 391751
 QY 264 GlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspValIleGlyThr 283
 DB 391752 GGAATTAATCTTTATTTGCTTGCCCAACCTGTTCTGCCAAGAAATTTGATGTAATCGTACA 391811
 QY 284 ValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspValSerIleIle 303
 DB 391812 GTAAATGCGCTAGAACACACGCTTGAAGATATTTATTACCAATGATGATCTATTATC 391871
 QY 304 GlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyValThrGlyGly 323
 DB 391872 GGTGTGTAGTGAATGGTCTCGCGAGGCACCTGCTCCGATCTCGCGCTAAACGGCGGT 391931
 QY 324 AsnLysLysSerGlyLeuTyrGluAspGlyValArg---LysAspArgLeuAspAsn 342
 DB 391932 AACAAAAAAGCGTTATTATCTTGACGGAGAACGCCAAAGAGCGTTTGTATACAGAA 391991
 QY 343 AspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAspGluAla 362
 DB 391992 GATATAGTGAACCAATTTAGAACCAAAATTCGTGCGAAAGTCGCACGACAGATCCAAAA 392051
 QY 363 ArgArgIle 365
 DB 392052 AACAGAAAT 392060

RESULT 6
 ABA92787
 ID ABA92787 standard; DNA; 640681 BP.
 XX
 AC ABA92787;
 XX
 DT 27-MAR-2002 (first entry)
 XX
 DE Buchnera sp. genomic DNA SEQ ID NO:1.
 XX
 KW Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
 KW circular; ds.
 XX
 OS Buchnera sp.
 XX
 PN JP2001292771-A.
 XX
 PD 23-OCT-2001.
 XX
 PF 07-APR-2000; 2000JP-0107160.
 XX
 PR 07-APR-2000; 2000JP-0107160.
 XX
 PA (RIKA) RIKAGAKU KENKYUSHO.
 XX
 DR WPI; 2002-126043/17.
 XX
 PT A genomic DNA of cockroach-symbiotic bacterium -
 XX

(MONS) MONSANTO TECHNOLOGY LLC.

Boronat A, Campos N, Rodriguez-concepcion M, Rohmer M, Seeman M;
Valentin HE, Venkatesh TV, Venkatramesh M;
WPI; 2002-227151/28.
P-PSDB; AAE19653.

gcpE nucleic acid which is an essential gene of the methyl-D-erythritol
phosphate pathway, encoding a fully defined GCPE protein which is
useful for increasing levels of tocopherol substrates in plants -
Claim 3; Page 117-119; 155pp; English.

The invention relates to gcpE nucleic acid molecule, an essential gene
of methyl-D-erythritol phosphate (MEP) pathway) that encodes rice,
Arabidopsis thaliana or Escherichia coli GCPE protein. gcpE is useful
for producing a transgenic plant such as Brassica campestris, B. napus,
canola, castor bean, coconut, cotton, crame, linseed, maize, mustard,
oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower,
or wheat with an increased isoprenoid (tocopherol) compound level. The
expression of GCPE protein in organisms increases the level of
tocopherol substrate such as isopentyl diphosphate and dimethylallyl
diphosphate biosynthesis. Transgenic organisms overexpressing GCPE
protein can nutritionally enhance food and feed sources. Overexpression
of GCPE protein in transgenic plant may provide tolerance to stresses
e.g., oxidative stress tolerance such as to oxygen or ozone, UV
tolerance, etc. gcpE may be used to obtain nucleic acid molecules from
the same species, and to obtain nucleic acid homologues. gcpE is also
used as or primers. The recombinant vectors are used in plant
transformation or transfection. gcpE an also act as markers capable of
detecting polymorphisms such as single nucleotide polymorphisms (SNPs).
gcpE is also used to determine the level or pattern of expression of
the protein. The present sequence is Escherichia coli gcpE gene.

SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Alignment Scores:

Pred. No.:	7.12e-184	Length:	1119
Score:	1866.00	Matches:	372
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	24	Gaps:	0

US-09-921-992-50 (1-372) x AAD31203 (1-1119)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
DB 1 ATGCATAACCGGCTCCATTTCACCTGAGAAATCAACAGTATTACGTTGGGAATGTG 60
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB 61 CCGATTGGCGATGCTGCCATCCGTCAGTCCATGACCAATACGGCTACGACAGAC 120
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
DB 121 GTCGAAGCAACGGTCAATCAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGFCCT 180
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAsn 80
DB 181 GTATCCGTACCGAGTATGACGGCGGAGAGCGTTCAACTCATCAACAGCAGGTTTAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
DB 241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTCGCTGAGAAAGTAGCGAATAC 300
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
DB 301 GCGCTCGATTGCTGCGTATTACCCCTGGCAATATCGTAAATCAAGAGCGTATTGCGCATG 360
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140

DB 361 GTGGTTGACTGTGCGCGCATAAAAACATTCCGATCCGATTGGCGTTAAACGCCGATCG 420
QY 141 IeuGluLysAspLeuGlnGlnLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
DB 421 CTGGAAAAAGATCTCAAGAAAAAGTATGCGCAACCGACGCCCGCAGCGGTTGCTGGAATCT 480
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
DB 481 GCATGGCTCATGTTGATCATCTCCATCGCCTGAACCTTCGATCAGTTCAAAGTCAGCGTG 540
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB 541 AAAGCGTCTGACGCTCTTCCTCGCTGTGAGTCTTATCGTTGCTGCGAAAACAGATCGAT 600
QY 201 GlnProLeuHisIeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220
DB 601 CAGCCGTTGTCATCTGGGATCAACCAAGCGGTGGTGGCGCAGCGGGCAGTAAATCC 660
QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
DB 661 GCATTGGTTTAGGCTCTGCTGCTGTAAGGCATCGGCGACACCGCTGCGGTATCGCTG 720
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLysLysSerLeuArgIle 260
DB 721 CGCGCGGATCCGTCGAAGAGATCAAGTCGGTTTCGATATTTTGAAATCGCTCGGTATC 780
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
DB 781 CGTTCGCGAGGATCAACTTCATCGCCTGCCGACCTGTTCCGCTCAGGAATTTGATGTT 840
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
DB 841 ATCGGTAGCTTTAAACGCTGGAGCAACGCTTGAAGATATCATCATCCGATGACGTT 900
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
DB 901 TCGATTATCGGCTGCTGCTGTAATGCCCGAGGCTGAGCGCTGTTTCTACACTCGCGTC 960
QY 321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
DB 961 ACCGGCGGCACACAGAAAAAGCGCCTCTATGAAGATGCGCTGCGCAAGACCGTCTGGAC 1020
QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
DB 1021 AACAACGATATATCGACCGCTGGAAGCACCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080
QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
DB 1081 GAAGCGCGTCGAATTGACGTTACGAGGTTGAANA 1116

RESULT 5
AAT42063
ID AAT42063 standard; DNA; 1830121 BP.
XX
AC AAT42063;
XX
DT 14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
KW Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; ds.
XX
OS Haemophilus influenzae.
XX
PN WO9633276-A1.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US05320.
XX

XX WPI: 2001-025196/03.
 DR P-PSDB; AAB45692.
 XX
 PT Incorporating gcpE and yfgB genes into viruses and cells, for
 PT increasing isoprenoid content and identifying e.g. antimicrobial
 PT agents, comprises using DNA sequences from bacteria or parasites
 XX
 XX Claim 3; Page 13-15; 36pp; German.
 XX
 CC This invention describes a novel method for incorporating gcpE and yfgB
 CC genes into viruses and cells for increasing isoprenoid content and
 CC identifying e.g. antimicrobial agents, comprises using DNA sequences (I)
 CC from the gcpE or yfgB genes of bacteria or parasites or DNA sequences (II)
 CC (II) which hybridize to the specified genes or encode a plastid protein
 CC with the same biological activity as those encoded by the genes. The
 CC invention also describes (1) plant cells containing (I) or (II); (2)
 CC transformed plant cells, and transgenic plants regenerated from them,
 CC that contain (I) or (II); (3) determining the enzymatic activity of a
 CC gcpE protein; or (4) screening compounds (A) that have antimycotic,
 CC antiparasitic or antiviral activity in humans or animals or antiviral,
 CC antiparasitic, fungicidal or herbicidal activity in plants. (I) and (II)
 CC are used: (i) to increase the isoprenoid levels in viruses and cells;
 CC (ii) for determining the enzymatic activity of gcpE and yfgB proteins;
 CC and (iii) to identify compounds that inhibit activity of gcpE, i.e.
 CC potential antibacterial, antimycotic, antiparasitic or antiviral agents
 CC for use in humans or animals, or antiviral, antiparasitic, fungicidal
 CC or herbicidal agents for agriculture.
 XX
 SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Alignment Scores:

Pred. No.: 7,12e-184 Length: 1119
 Score: 1866.00 Matches: 372
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 22 Gaps: 0

US-09-921-992-50 (1-372) x AAC82653 (1-1119)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
 DB 1 ATGCATAACCCAGGCTCCAAATTCACCTAGAGAAATCAACAGTATTACGTTGGGAATGTG 60
 QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40
 DB 61 CCGATTGGCGATGGTGTCCCATCGCGGTACAGTCCATGACCAATACGCGTACGACAGAC 120
 QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
 DB 121 GTCAAGCAACGGTCAATCAATCAAGGCGCTGGAACGGCTTGGCGTGTATATCGTCCGT 180
 QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
 DB 181 GTATCCGTACCGAGTGGACGGCGGAGAGCGTTCAAACTCATCAACAGCAGGTTAAC 240
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
 DB 241 GTGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTCGCTGAAAGTAGCGGAATAC 300
 QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
 DB 301 GGGCTCGAATGTCTGGCTATTAACCTTGGCAATATCGTAAATGAAGCGGTATTCGCATG 360
 QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
 DB 361 GTGGTTGACTGTGGCGCGATNAACATTCGATCCGATCCGATTGGCGTTAAACGCGGATCG 420
 QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
 DB 421 CTGGAAGAAAGATCTGCAAGAAAAAGTATGGCGAAACCGACCGCGCGGCTGTGCTGGAATCT 480

QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
 DB 481 GCATGGCTCATGTTGATCATCTCGATCGCTGAATTCGATCAGTTCAAAGTCAGCGTG 540
 QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
 DB 541 AAAGCGCTCGACGCTCTTCCTCGCTGTAGTCTTATCGTTCTGCGCAAAACAGATCGAT 600
 QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220
 DB 601 CAGCCGTTGTCATCTGGGATACACCAAGCCGTTGGTCCGCGAGCGGCGAGTAAATCC 660
 QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
 DB 661 GCATTTGGTTTAGTCTGCTGCTGCTGAAGCATCGGCGACACGCTGCGGTATCGCTG 720
 QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
 DB 721 GCGGCGCATCCGCTCGAAGAGATCAAGTTCGGTTCGATATTTGAAATCGCTGCGTATC 780
 QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
 DB 781 CATTCCGAGGATCAACTTCATCGCTGCCGACCTGTTCGCGTCAGGAATTTGATGTT 840
 QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
 DB 841 ATCGGTACGTTAAACGCGCTGGAGCAACGCTTGAAGATATCATCACTCCGATGACGTT 900
 QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
 DB 901 TCGATTATCGGCTCGTGTGAATGCCAGGTGAGCGCTGTTCTACACTCGCGCTC 960
 QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
 DB 961 ACCGCGGCAACAAAGAAAGCGCTCTATGAAGATGGCGTGGCAAAACCGCTCTGGAC 1020
 QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
 DB 1021 AACACGATATGATCGACCATCGAGCTGGAGCACGATTCGTCGAAGCCAGTCAGCTGGAC 1080
 QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
 DB 1081 GAAGCGCTCGAATTCAGCTTCAGCAGGTGTAAGAAA 1116
 RESULT 4
 AAD31203
 ID AAD31203 standard; DNA; 1119 BP.
 XX
 AC AAD31203;
 DT 31-MAY-2002 (first entry)
 XX
 DE Escherichia coli gcpE gene.
 XX gcpE gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate;
 KW transgenic plant; isoprenoid compound; tocopherol; isopentyl diphosphate;
 KW food; feed source; transfection; single nucleotide polymorphism; SNP;
 KW oxidative stress tolerance; UV tolerance; transformation; plant; ds.
 XX Escherichia coli.
 OS
 FH Key Location/Qualifiers
 FT CDS 1..1119
 FT /*tag= a
 FT /product= "Escherichia coli GCPE protein"
 XX
 XX WO200212478-A2.
 XX
 XX 14-FEB-2002.
 PD
 XX
 XX 06-AUG-2001; 2001WO-US24335.
 PF
 XX
 XX 07-AUG-2000; 2000US-223483P.
 PR

DR WPI; 2000-640125/62.

XX Identifying antagonists of the expression of gene encoding bacterial

PT growth polypeptide useful for treating bacterial infections or

PT diseases, by evaluating transcription of the gene in the presence of

PT test molecule -

XX Claim 1; Page 25; 55pp; English.

PS The present sequence is that of the Escherichia coli FUN gene

XX gcpE, which encodes a protein that is essential for bacterial

CC growth or survival. gcpE is 1 of 22 E. coli genes (see

CC AAA8692-713) identified as being essential (there is no deletion

CC genotype). These 22 genes fulfil criteria for being attractive

CC antibacterial targets: hypothetical open reading frames coding for

CC essential functions (mutation is lethal for growth in rich media);

CC broad conservation (orthologues are present in a wide range of

CC bacteria, including Haemophilus influenzae, Streptococcus pneumoniae,

CC Helicobacter pylori and Borrelia burgdorferi); and low toxicity

CC potential in higher organisms (mostly no orthologues were identified

CC in Saccharomyces cerevisiae). An antagonist or inhibitor of the

CC expression of an essential gene or of its function provides the key

CC for antibacterial therapy. The invention provides methods for

CC identifying such antagonists or inhibitors. These involve

CC contacting a bacterial cell comprising an essential gene with a

CC candidate antagonist or inhibitor, and testing whether contact leads

CC to cell growth inhibition and/or cell death. The method allows the

CC development of new broad spectrum antibiotics. A conditional mutant

CC of an essential gene can be used to induce a lethal phenotype in

CC bacteria for the analysis of surrogate markers.

XX

SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Alignment Scores:

Pred. No.:	7.12e-184	Length:	1119
Score:	1866.00	Matches:	372
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	21	Gaps:	0

US-09-921-992-50 (1-372) x AAA88705 (1-1119)

QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20

DB 1 ATGCNTAACCGAGCTCAATTCACGTAGAAATCAACAGTATTACGTTGGGAATGG 60

QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrThrAsp 40

DB 61 CGGATTGGCGATGGTGCTCCCATCGCGTACAGTCCATGACCAATACGCGTAGCAGAC 120

QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60

DB 121 GTCGAAGCAACGGTCAATCAAAATCAAGGCGCTGGAACGGCTTGGCGCTGATATCGTCCGT 180

QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnValAlaen 80

DB 181 GATTCGTTACCGACGATGACGGCGGAGAGCGGTTCAAACTCATCAACAGCAGGTTAAC 240

QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100

DB 241 GTGCCGCTGGTGCTGACATCCACTTCGATTCGATTCGCTGAAAGTAGCGGAATAC 300

QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluArgIleArgMet 120

DB 301 GCGCTCGATTGTCTGCGTATTACCTCGCAATATCGTAATGAAGAGCGTATTTCGATG 360

QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140

DB 361 GTGGTTGACTGTGCGCGCGGATATAAAACATTCGATCCGATTGGCGTTAACGCCGATCG 420

QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160

DB 421 CTGGAAGAAAGATCTGCAAGAAAGTATGCGCAACCGACCGCGCGTTCGTGGAATCT 480

QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180

DB 481 GCATGCGTCAATGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAAGTCAGCGTG 540

QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200

DB 541 AAAGCGTCTGACGCTCTCTCTCGCTGTTGAGTCTTATCGTTTCTGCGCAACACAGATCGAT 600

QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyAlaArgSerGlyAlaValLysSer 220

DB 601 CAGCCCTTCATCTGGGATCACCGAAGCGGTGGTGGCGCGAGCGGCGAGTAAATCC 660

QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240

DB 661 GCATTGGTTTGGTCTGCTGCTGAAGGATCGCGACACGCTGCGCGTATCGCTG 720

QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260

DB 721 GCGGCGATCCGTCGAGAGATCAAGTCGGTTTCGATATTTTGAATCGCTGCGTATC 780

QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280

DB 781 CGTTCGCGAGGATCAACTTCATCGCTCGCGACCTGTTCCGCTCAGGAATTTGATGTT 840

QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300

DB 841 ATCGGTACGGTTAACGCGCTGGAGCAACGCTTGAAGATATCATCACTCCGATGGAAGTT 900

QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320

DB 901 TCGATTATCGGTCGCTGCTGTAATGCGCCAGGTGAGCGCTGTTTCTACACTCGCGCTC 960

QY 321 ThrGlyGlyAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340

DB 961 ACCGCGCGCAACAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGACCGTCTGGAC 1020

QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360

DB 1021 AACAAAGATATGATCGACCGCTGGAAGCACCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372

DB 1081 GAAGCGCTCGAATTGACGTTTCAGCGGTTGAAAAA 1116

RESULT 3

AAC82653

ID AAC82653 standard; DNA; 1119 BP.

XX

AC AAC82653;

XX

DT 15-MAR-2001 (first entry)

XX

DE E. coli gcpE DNA.

XX

KW Isoprenoid; gcpE; YfgB; antimicrobial; transgenic plant; agriculture;

KW antimycotic; antiparasitic; antiviral; fungicidal; herbicidal; ds.

XX

OS Escherichia coli.

XX

PN WO200072022-A1.

XX

PD 30-NOV-2000.

XX

PF 20-MAY-2000; 2000WO-EP04592.

XX

PR 21-MAY-1999; 99DE-1023567.

PR 21-MAY-1999; 99DE-1023568.

XX

PA (JOMA/) JOMAA H.

XX

PI Jomaa H;

XX WO20061793-A2.
PN 19-OCT-2000.
XX 07-APR-2000; 2000WO-EP03135.
PF 09-APR-1999; 99EP-0107031.
PR 04-FEB-2000; 2000EP-0102111.
XX (GPCB-) GPC BIOTECH AG.
PA Loferer H, Jacobi A;
XX WPI; 2000-697048/67.
DR
XX
XX
PT Identifying antibacterial compounds, comprises identifying an
PT antagonist or inhibitor of the expression of a gene encoding a
PT polypeptide essential for bacterial growth or survival -
XX
XX Claim 1; Fig 1; 75pp; English.
XX
XX The present invention relates to antagonists and inhibitors of 24
CC bacterial genes and proteins. The proteins are thought to be essential
CC for growth in several species of bacteria (including *S. pneumoniae*, *B.*
CC *burgdorferi*, *H. influenza* and *H. pylori*). The proteins and coding
CC sequences shown in the specification can be used to identify antagonists
CC and inhibitors which can be used in disease treatment and pesticides. In
CC particular, they can be used against *M. tuberculosis*. The present
CC sequence is one of the genes of the invention.
XX
XX Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;
SQ
Alignment Scores:
Pred. No.: 7,12e-184 Length: 1119
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0
US-09-921-992-50 (1-372) x AAA95478 (1-1119)
QY 1 MetHisAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
DB 1 ATGCATAACCGAGCTCCAAATCAACGTAGAAATCAACACGTAATTCAGTTGGGAATG 60
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB 61 CCGATTGGCGATGGTCTCCATCCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
DB 121 GTCGAAGCAACGGTCAATCAAAATCAAGCGCTGGAACGGCTGGCGCTGATATCGTCCGT 180
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheIleLeuIleLysGlnGlnValAsn 80
DB 181 GTATCCGTACCGACGATGGACGGCGGAGACGCTTCAAACTCATCAACAGCAGGTAAAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
DB 241 GTGCCCGTGGTGGCTGACATCCACTTCGACTATCGCATTCGGCTGAAAGTAGCGGAATAC 300
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
DB 301 GGCGTCCGATTGTCTGGCTATTAAACCCCTGGCAATATCGTAAATGAAGACGCTATTCGCATG 360
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAsnAlaGlySer 140
DB 361 GTGGTTGACTGTGGCGCGATAAACATTCCTCGATTCGCTATTCGCGTAAACGCGGATCG 420
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160

DB 421 CTGGAAGAAAGATCTGCAAGAAAGATATGGCGAACCGCCGCGCGCTTGTGGAATCT 480
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheIleValSerVal 180
DB 481 GCCATGCGGTCAATGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB 541 AAAGCGTCTGAGCTCTTCTCGCTGTTGAGTCTTATCGTTGCTGCAAAAACAGATCGAT 600
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
DB 601 CAGCGCTTCATCTGGGATCACCGAAGCGGTGGTGGCGCGCGGCGGCAATAATCC 660
QY 221 AlaIleGlyLeuGlyLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
DB 661 GCCATTGGTTTAGGTCTGCTGCTGCTGAAGCATCGGCGACACGCTGGCGGTATCGCTG 720
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB 721 GCGGCGGATCCGTCGAAGAGATCAAGTCGTTTCGATATTTTGAATCGCTGCGTATC 780
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280
DB 781 CGTTCGCGAGGATCAACTTCATCGCTGCCGACCTGTTCGCGTCAGGAATTTGATGTT 840
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
DB 841 ATCGGTACGGTTAAACGCGCTGAGCAACGCTGGAAGATATCATCACTCCGATGGACGTT 900
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
DB 901 TCGATTATCGGCTGGTGGTGAATGCCAGGTGAGCGCTGTTTCTACACTCGCGCTC 960
QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
DB 961 ACCGGCGGCAACAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGACCGTCTGGAC 1020
QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
DB 1021 AACAACGATATATGATGACCGCTGGAAGCAGCATTCGTGGCAAGCCAGTCAGCTGGAC 1080
QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
DB 1081 GAAGCGCTCGAATTGACGTTTCAGCGTTGAAAAA 1116
RESULT 2
AAA88705
ID AAA88705 standard; DNA; 1119 BP.
XX
AC AAA88705;
XX
XX 05-FEB-2001 (first entry)
DT
XX
DE E. coli FUN essential gene gcpe.
XX
KW FUN gene; gcpe gene; essential gene; antibacterial; antibiotic;
XX screening; infection; therapy; antagonist; surrogate marker; ds.
OS Escherichia coli.
XX
XX EP1043403-A1.
XX
PD 11-OCT-2000.
XX
XX 09-APR-1999; 99EP-0107031.
XX
XX 09-APR-1999; 99EP-0107031.
XX (GPCB-) GPC GENOME PHARM CORP AG.
XX
XX Not given;
XX

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: November 23, 2003, 17:17:14 ; Search time 336 Seconds
(without alignments)
2988.664 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866

Sequence: 1 MNQAPIQRKSTRIYGVN.....RAKASQLDEARRIDVQQVEK 372

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame_p2n.model -DRV=xlh
-Q=/cgn2_1/USPTO.spool/US09921992/runat_21112003_162545_18397/app_query.fasta_1.519
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -NORM=ext -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09921992.@CGN_1_0@runat_21112003_162545_18397 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DRV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Databases :

N Geneseq 19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1866	100.0	1119	21	AAA95478	E. coli essential
2	1866	100.0	1119	21	AAA88705	E. coli FUN essent
3	1866	100.0	1119	22	AA832653	E. coli gcpE DNA.
4	1866	100.0	1119	24	AAD31203	Escherichia coli g
5	1603.5	85.9	1830121	17	AAT42063	Haemophilus influ
6	1282	68.7	640681	24	ABA92787	Buchnera sp. genom
7	1166	62.5	33140	22	AAF28536	Genomic fragment #
8	868	46.5	1107	24	ABQ67966	Listeria monocytog
9	868	46.5	1107	24	ABQ63993	Listeria monocytog
10	868	46.5	2944528	24	ABA03041	Listeria monocytog
11	861.5	46.2	6157	24	ABQ70939	Listeria lichenifo
12	824	44.1	1083	24	ABK74401	Bifidobacterium lo
13	814.5	43.6	349980	24	ABQ81842	Listeria monocytog
14	812	43.5	994	24	ABQ68324	H. pylori GHPO 16
15	794	42.6	1180	19	AA14017	Propionibacterium
16	785.5	42.1	35829	23	AAS59573	C glutamicum codin
17	761	40.8	1134	22	AAH67170	C. glutamicum deri
18	761	40.8	1179	25	ACA00649	C. glutamicum codin
19	761	40.8	349980	22	AAH68530	C. glutamicum codin
20	761	40.8	349980	22	AAH68531	Mycobacterium tube
21	751.5	40.3	1161	25	ABZ71126	Mycobacterium tube
22	751.5	40.3	1164	22	AAH52056	Mycobacterium tube
23	751.5	40.3	4403765	22	AAI99683	Mycobacterium tube
24	751.5	40.3	4411529	22	AAI99682	Mycobacterium tube
25	682	36.5	5484	23	AA573006	DNA encoding novel
26	534	28.6	1263	25	ABZ33756	N. gonorrhoeae nuc
27	527	28.2	60873	21	AAA81469	N. meningitidis pa
28	527	28.2	349980	21	AAF21610	Neisseria meningit
29	527	28.2	1437668	21	AAA81490	N. meningitidis B
30	510	27.3	1857	23	AA590066	DNA encoding novel
31	490	26.3	1038602	20	AAZ01425	Complete genome se
32	470.5	25.2	273254	21	AAC81914	Chlamydia pneumoni
33	470.5	25.2	1230025	20	AA391990	Nucleotide sequenc
34	462	24.8	2520	24	AAC31201	Arabidopsis thalia
35	460	24.7	2109	22	AA832654	P. falciparum gcpE
36	446	23.9	3147	21	AA829996	P. falciparum gcpE
37	428.5	23.0	507	21	AACT5580	Human OREX ORF135
38	428.5	23.0	507	24	ABN76496	Human ORF1443 cDNA
39	383	20.5	435	24	ABX66784	Helicobacter pylor
40	339	18.2	670	24	AAD31221	Zea mays partial g
41	326	17.5	596	24	AAD31222	Zea mays partial g
42	321	17.2	584	24	AAD31220	Zea mays partial g
43	300	16.1	4467	10	AA92428	Sequence encoding
44	292	15.6	33675	24	AAD31202	Oryza sativa gcpE
45	277.5	14.9	2535	20	AA20674	Polynucleotide seq

ALIGNMENTS

RESULT 1
AAA95478
ID AAA95478 standard; DNA; 1119 BP.
XX
AC AAA95478;
XX
DT 27-FEB-2001 (first entry)
XX
DE E. coli essential gene gcpE.
XX
KW Bacterial growth; inhibitor; yggB; yfhC; yacE; ychB; vejD; yrfI;
KW yggJ; yjeE; yjaO; yrdC; yhcC; ygbC; ybeC; ygpE; kdsB; pfs; yeaJ;
KW B108; yeaA; yagP; B193; yidd; yceG; yjbc; antibacterial compound;
KW H. influenza; S. pneumoniae; H. pylori; B. burgdorferi; pesticide;
KW M. tuberculosis; antibiotic; ds.
XX
OS Escherichia coli.

THIS PAGE BLANK (USPTO)

Qy 206 CAGAAAGCGTTCAAACTCATCAA 227
| | | | |
Db 658 CTGATGCCTGCTTTGAGATCAA 679
| | | | |

Search completed: November 23, 2003, 15:53:54
Job time : 2756 secs

Db 330 CCACGGCGCCATGATGGTGGCGACCTTCGAGAAGATCCGTGTGAACCCCGGCAACT 389
QY 335 TCGGTAATG 343
Db 390 TCCTGACG 398

RESULT 14
AV641368
LOCUS
DEFINITION AV641368 Chlamydomonas reinhardtii 5% CO2 linear EST 15-DEC-2000
CDNA clone HCL032409_r 5', mRNA sequence.
ACCESSION AV641368
VERSION AV641368
KEYWORDS 1 GI:10784696
SOURCE Chlamydomonas reinhardtii
ORGANISM Chlamydomonas reinhardtii
REFERENCE AV641368.1 Chlamydomonas reinhardtii
AUTHORS Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohya, K., Nakamura, Y., and Tabata, S.
TITLE Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii
JOURNAL DNA Res. 7 (5), 305-307 (2000)
MEDLINE 20539644
PubMed 11089912
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

FEATURES
1..377
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="HCL032409_r"
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from cells cultured in a medium with bubbling air containing 5% carbon dioxide"
BASE COUNT 89 a 118 c 116 g 54 t
ORIGIN
Query Match 5.5%; Score 62; DB 9; Length 377;
Best Local Similarity 55.6%; Pred. No. 8.7e-08;
Matches 150; Conservative 0; Mismatches 105; Indels 15; Gaps 1;
QY 50 TTGGGAATGCGGATGGCGATGGTGTCTCCATCGCGTACAGTCCATGACCAATACGC 109
Db 102 TCGCAAGTGGCGGGTGGCGAGCAGCAGCGATCGCCCTTCAGACATACCAACACCG 161
QY 110 GTACGACAGACGTGCAAGCAACGGTCAATCAATCAAGGCGCTGGAACGGTGGCGCTG 169
Db 162 ACACCGCATGTGCGAGTGCAGTGCAGCGACGAGTGAAGAGTGGCGGAGCGCGCGG 221
QY 170 ATATCGTCGTGTATCGTATCCGACGATGACGCGGAGAGCGTTCA----- 217
Db 222 ACATCGTCGCATCACAGTGCAGGCAAGAGGAGGCTGAGGCGTGCATGAAGATTGCG 281
QY 218 ---AATCATCAACAGCAGGTAACTGCGCGTGGTGGTGCATCCATCCATTCGATTC 274
Db 282 AGCAGTGTTCAGAGCAAGTACGACGTGCCCTGTGGCGGACATCCACTTCACGCCCA 341
QY 275 GCATTGGCTGAAAGTACCGGAATACGGCG 304
Db 342 CGGTGGCCATGATGGTGGCGGACGCCTTCG 371

RESULT 15

CD424926
LOCUS
DEFINITION SA1_9 D05_g1 A002 Salicylic acid-treated seedlings Sorghum bicolor
CDNA Clone SA1_9_D05_A002 5', mRNA sequence.
ACCESSION CD424926
VERSION CD424926.1
KEYWORDS 1 GI:31331189
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.
AUTHORS 1 (bases 1 to 688)
Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Salzman, R., Chua Tan, N., Gonzalez, M., Lane, S., Miller, V., Nanda, P., Olaseinde, O., Eastman, A. and Pratt, L.H.
TITLE An EST database from Sorghum: salicylic acid-treated seedlings
JOURNAL Unpublished
COMMENT Other ESTs: SA1_9 D05.bl A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.
Seq primer: Sugs (CTTCTGCTCTAAAGCTGG).
1..688
Location/Qualifiers
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="SA1_9 D05_A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/note="Vector: pME18S-FU3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from seedlings grown in hydroponic culture. At 8 days of age, medium was supplemented with 1 mM salicylic acid (SA). Roots and shoots were harvested after 27 and 72 hr and material from both time points was combined prior to RNA isolation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FU3 vector (5'-prime DraIII site is CACTGTGTG, 3'-prime DraIII site is CACATGTG). XhoI excises the cDNA insert."
BASE COUNT 174 a 172 c 205 g 137 t
ORIGIN
Query Match 5.2%; Score 58; DB 14; Length 688;
Best Local Similarity 55.4%; Pred. No. 2e-06;
Matches 112; Conservative 0; Mismatches 90; Indels 0; Gaps 0;
QY 26 GTAGAAATCAACACGCTATTTACGTTGGGAATGTCGATGTCGATGTCCTCCATCG 85
Db 478 GGAGGAAACTCGAAGTGTGATGGTGGGAATGTGCCACTTGGCAGTGATCATCCATAA 537
QY 86 CGGTACAGTCCATGACCAATACGCTAGCAGAGAGCTGGAAGCAACGGTCAATCAATCA 145
Db 538 GGATTCACCAACCATGACGCTTCAGATACCAAGAGTGTTCAGAAACAGTAGAGGAGTGA 597
QY 146 AGCGCTGGAACGCGTTGGCGGTGATATCGTCCGTGTATCCGTACCGACGATGGACGG 205
Db 598 TGAGGATACGAGATAAAGGAGCTGATATTGTTAGATAACCGTTTCAGGTGAGGAAGAG 657

Matches	175;	Conservative	0;	Mismatches	116;	Indels	18;	Gaps	2;
Qy	50	TTGGGAATGTC	CGGATTTGGCGATGGTGTCTCCCATCGCGGTACAGTCCATGACCAATACGC	109					
Db	99	TCGCAAGGTGCGGTGGCAGCCAGCACCGCATCGCCCTTCAGACCATGACCCACCGG	158						
Qy	110	GTACGACAGACGTCGAAGCAACGGTCAATCAAAATCAAGGCGCTGGAAAGCGTTGGCGGTG	169						
Db	159	ACATCCGCAATGTGTGAGCTGACCGTGGACCGACAGGTGAAGAAAGTGC	218						
Qy	170	ATATCGTCCGTTGATTCGTACCGACGATGGAGCGCGGCAAGACGGTTCA	217						
Db	219	ACATCGTCCGATCACAGTGCAGGGCAAGAGAGGCTGAGCGGTGATGAAGATTCGCG	278						
Qy	218	--AACTCATCAAAACAGCAGGTTAAAGTGC	274						
Db	279	AGCAGCTGTTCAAGGACAAGTACGACGTGCGCCCTGGTGGCGCATCCACTTCGAGCCCA	338						
Qy	275	GCATTGGCGTGAAGTAGCGGAATACGGCGTCGATTGTCTCGGTATTAAACCGCTGGCAATA	334						
Db	339	CGGTGGCCATGATGTTGGCGG--ACGCTTCGAGAGAAGATCCGTGTGAAACCCCGGCAACT	395						
Qy	335	TCGGAATG	343						
Db	396	TCGCTGACG	404						

RESULT 8	
LOCUS	AV626792
DEFINITION	502 bp mRNA linear EST 15-DEC-2000 AV626792 Chlamydomonas reinhardtii 5' to 0.04% CO2 Chlamydomonas reinhardtii cDNA clone LCL015e03_r 5', mRNA sequence.
ACCESSION	AV626792
VERSION	AV626792.1 GI:10789072
KEYWORDS	EST.
SOURCE	Chlamydomonas reinhardtii
ORGANISM	Chlamydomonas reinhardtii Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales; Chlamydomonadales; Chlamydomonas.
REFERENCE	1 (bases 1 to 502)
AUTHORS	Asamizu, E., Miura, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y. and Tabata, S.
TITLE	Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii
JOURNAL	DNA Res. 7 (5), 305-307 (2000)
MEDLINE	20539644
PUBMED	11089912
COMMENT	Contact: Erika Asamizu The First Laboratory for Plant Gene Research Kazusa DNA Research Institute Yama 1532-3, Kisarazu, Chiba 292-0812, Japan Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/

```

FEATURES             location/Qualifiers
source               1..502
                     /organism="Chlamydomonas reinhardtii"
                     /mol_type="mRNA"
                     /strain="C9"
                     /db_xref="taxon:3055"
                     /clone="LCLO15e03_r"
                     /clone_lib="Chlamydomonas reinhardtii 5% to 0.04% CO2"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; The cDNA library was constructed from cells cultured
                     in a carbon stress acclimatized condition in which carbon
                     dioxide concentration in the bubbling gas was changed from
                     5% to 0.04%"
BASE COUNT          114 a   157 g   153 G   78 t
ORIGIN
Query Match          6.0%; Score 67.4; DB 9; Length 502;
Best Local Similarity 56.8%; Pred. No. 2.2e-09;
Matches 175; Conservative 0; Mismatches 116; Indels 18; Gaps 2;

QY 50 TTGGGAATGTGCGGATTGGCGATGGTGGTCTCCCATCGCGTACAGTCATGACCAATACGC 109

```

Db	107	TCGCGAAGTGC	CGGTGGCAGCAGCACCGCATCGCCTTTCAGACCATGACCACCGG	166
Qy	110	GTACGACAGACGTCGAAGCAACGGTCAATCAAAATCAAGGCGCTGGAACGCGTTGGCGCTG	169	
Db	167	ACACCCGCATGTGCAGCTACCGCTGGACACAGGTGAAGAAGTCTCGCGGAGCCCGGCGCGG	226	
Qy	170	ATATCGTCCGTGTATCCGTACCGACGATGACGCGGACGAAGCGTTCA	217	
Db	227	ACATCGTCCGCATCACAGTGCAGGCAAGAAGAGAGCTGAGGCGTGCATGAAGATTTCGCG	286	
Qy	218	---AATCATCAACAGCAGGTTAACTGTCGCGCTGGTGGCTGACATCCACTTCGACTATC	274	
Db	287	AGCAGCTGTTTCAAGGACAAGTACGAGCTGCCCTCGTGGCGCAGCATCCACTTCCAGCCCA	346	
Qy	275	GCATTGCGCTGAAAGTAGCGGAATACGGCGTGCAGTGTCTGCGTATTTAACCCCTGGCAATA	334	
Db	347	CGGTGGCCATGATGTTGGCGG---ACGCTTCGAGAAGATCCGTGTGAACCCCGGCAACT	403	
Qy	335	TCGCTAATG	343	
Db	404	TCGCTGACG	412	

RESULT 9

AV644476

LOCUS

DEFINITION

AV644476 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii cDNA clone HCL089b10_r 5', mRNA sequence.

ACCESSION

AV644476

VERSION

AV644476.1 GI:10787804

KEYWORDS

EST.

SOURCE

Chlamydomonas reinhardtii

ORGANISM

Chlamydomonas reinhardtii

REFERENCE

AUTHORS

Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K., Nakamura, Y. and Tabata, S.

TITLE

Generation of expressed sequence tags from low-CO2 and high-CO2 adapted cells of Chlamydomonas reinhardtii

JOURNAL

DNA Res.

MEDLINE

7 (5), 305-307 (2000)

PUBMED

20539644

COMMENT

11089912

Contact: Erika Asamizu

The First Laboratory for Plant Gene Research

Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 282-0812, Japan

Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.

Location/Qualifiers

1..512

FEATURES

source

```

email: asanuma@kaxa.co.jp; URL: http://www.kaxa.co.jp/eu/print/
Location/Qualifiers
1. 512
/organism="Chlamydomonas reinhardtii"
/mol_type="mrna"
/strain="C9"
/db_xref="taxon:3055"
/clone="HCL089b10_r"
/clone_lib="Chlamydomonas reinhardtii 5% CO2"
note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"
BASE COUNT 118 a 159 c 158 g 77 t
ORIGIN
Query Match 6.0%; Score 67.4; DB 9; Length 512;
Best Local Similarity 56.6%; Pred. No. 2.2e-09;
Matches 175; Conservative 0; Mismatch 116; Indels 18; Gaps 2;

Qy 50 TTGGGATGTCGGATGGCGATGGTCTCCCATGCCCGTCAGTCACCAATACGC 109
Db 96 TCGGCAAGGTGCGGTGGCGACCCAGCACCCGATGCCCTTCAGACCATGACCAACCG 155
Qy 110 GTACGACAGACGTCGAAGCAACGGGTCAATCAAAATCAAGCGCTGGAAACGGCTTTGGCGCTG 169

```

```
FEATURES
source
Class: shotgun.
Location/Qualifiers
1..1268
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164.970"
/clone_lib="pacs2-164"
/note="clinical isolate 2-164 Whole genomic shotgun
library."
BASE COUNT 257 a 386 c 351 g 273 t 1 others
ORIGIN
Query Match 6.1%; Score 68.2; DB 29; Length 1268;
Best Local Similarity 58.8%; Pred. No. 2e-09;
Matches 137; Conservative 0; Mismatches 93; Indels 3; Gaps 1;
QY 853 AACGCGCTGGAGCAAGCCCTGGAAGATATCATCTCCGATGACCGTTTCGATTATCGGC 912
Db 752 AAGGACTGGAAGGGCGCTGGAGACCTCTGATCCGATGACCGTGGCGGTGATCGGT 693
QY 913 TGGTGGTGAATGCCAGGTAGGCGCTGGTTTCTACATCGGCGTACCGGCGGCAAC 972
Db 692 TGGTGGTCAACGGTCCGTGCGAAGCAAGAGGCGCATGTGCGGCTCTACCGGCTGCACT 633
QY 973 AGAAGAGCGGCTCTATGAGATGCGTGGCGCAAGACCGTCTGGACACACGATATG 1032
Db 632 CGGAA---CCTGGTGTATATCGACGGCAAGCCGTGCGAGAACTGACCAACGACACCTG 576
QY 1033 ATGACGAGCTGGAAGCAGCATTCTGCGAAGCCAGTCAGTGGACGCAAGC 1085
Db 575 GTGAGCAGCTGGAACGGTGTATCCCGACAGCGCGCCGAGAGCGCCGAGGC 523
RESULT 6
AV644003 448 bp mRNA linear EST 15-DEC-2000
LOCUS
DEFINITION
AV644003 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
CDNA clone HCL080a02_f 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Chlamydomonas reinhardtii
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE
1 (bases 1 to 448)
AUTHORS
Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y. and Tabata, S.
TITLE
Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL
MEDLINE
PUBMED
DNA Res. 7 (5), 305-307 (2000)
COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..448
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="HCL080a02_r"
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"
BASE COUNT 106 a 142 c 132 g 68 t
ORIGIN
Query Match 6.0%; Score 67.4; DB 9; Length 482;
Best Local Similarity 56.6%; Pred. No. 2.1e-09;
Matches 176; Conservative 0; Mismatches 117; Indels 18; Gaps 2;
QY 50 TTGGGAATGTGCCGATTTGGCGATGGTGTCTCCATCCCGTACAGTCCATGACCAATACGC 109
Db 97 TCGGCAAGGTGCGGTGGGAGCAGCCGATCGCCCTTCAGACCATGACCAACCG 156
QY 110 GTACACAGAGCTCGAAGCAACGGTCAATCAATCAAGGCGCTGGACCGTTGGCGCTG 169
Db 157 ACACCCGCAATGTGCAGCTGACCGTGGACCGTGAAGAAGTCCGCGGACCGCGCGG 216
QY 170 ATATCGTCCGTATCTCCGTACCGACGATGGAGCGCGCAGACGCTTCA----- 217
Db 217 ACATCGTCCGATCAGTGGAGGCGAGAGAGGCTGAGCGGTGCATGAAGATTCGGC 276
QY 218 ---AACTCATCAACAGCAGGTTAAAGTCCGCTGGTGGCTGACATCCACTTCGACTATC 274
Db 277 AGCAGCTGTTCAAGGACAAGTACGACGTGCCCTGGTGGCGACATCCACTTCCAGCCCA 336
QY 275 GCATTGCGCTGAAGTAGCGGAATAGCGGTGCGATTGTCTGCGTATTACCTGGCAATA 334
Db 337 CGGTGGCCATGATGTGTGGCGG---ACGCTTCGAGAAGATCCGTGTGAACCCCGGCACT 393
QY 335 TCGGTAATGAA 345
Db 394 TCGCTGACGGA 404
RESULT 7
AV642898 482 bp mRNA linear EST 15-DEC-2000
LOCUS
DEFINITION
AV642898 Chlamydomonas reinhardtii 5% CO2 Chlamydomonas reinhardtii
CDNA clone HCL060b04_r 5', mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Chlamydomonas reinhardtii
Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.
REFERENCE
1 (bases 1 to 482)
AUTHORS
Asamizu, E., Miura, K., Kucho, K., Inoue, Y., Fukuzawa, H., Ohyama, K.,
Nakamura, Y. and Tabata, S.
TITLE
Generation of expressed sequence tags from low-CO2 and high-CO2
adapted cells of Chlamydomonas reinhardtii
JOURNAL
MEDLINE
PUBMED
DNA Res. 7 (5), 305-307 (2000)
COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
FEATURES
source
1..482
/organism="Chlamydomonas reinhardtii"
/mol_type="mRNA"
/strain="C9"
/db_xref="taxon:3055"
/clone="HCL060b04_r"
/note="Vector: pBluescriptII SK-; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from cells cultured
in a medium with bubbling air containing 5% carbon
dioxide"
BASE COUNT 112 a 150 c 147 g 73 t
ORIGIN
Query Match 6.0%; Score 67.4; DB 9; Length 482;
Best Local Similarity 56.6%; Pred. No. 2.1e-09;
```

KEYWORDS GSS.
SOURCE Spiroplasma kunkelii
ORGANISM Spiroplasma kunkelii
Bacteria; Firmicutes; Mollicutes; Entomoplasmatales;
Spiroplasmataceae; Spiroplasma.
REFERENCE 1 (bases 1 to 720)
AUTHORS Hogenhout, S.A.
TITLE Genomic sequences from Spiroplasma kunkelii strain M2
JOURNAL Unpublished
COMMENT Contact: Hogenhout SA
Department of Entomology
The Ohio State University-OARDC
120 Thorne Hall, 1680 Madison Avenue, Wooster, OH 44691, USA
Tel: 330 263 3730
Fax: 330 263 3686
Email: hogenhout.1@osu.edu
Class: HindIII fragments.
Location/Qualifiers
1. .720
/organism="Spiroplasma kunkelii"
/mol_type="genomic DNA"
/strain="M2"
/db_xref="taxon:47834"
/clone="PH_05.x"
/clone_lib="Spiroplasma kunkelii H"
BASE COUNT 255 a 96 c 132 g 233 t 4 others
ORIGIN
Query Match 7.3%; Score 81.6; DB 28; Length 720;
Best Local Similarity 53.4%; Pred. No. 1.1e-13;
Matches 191; Conservative 0; Mismatches 166; Indels 1; Gaps 1;
QY 572 CTTATCGTTTGGCAACACATCGATCAGCGGTGGCATCTGGGGATCCGAGCGG 631
Db 35 CTTATACCTAGCTAGTAAGAATGAATATCCCTTCATCTAGGAT-GCTGAAGCTG 93
QY 632 GTGGTGGCGGAGCGGGGAGTAAATCCGCCATTTGGTTAGTCTGTCTGTCTGAAG 691
Db 94 GTAGTCATCATACCGGAACANNAATCATGTAGTGTGTGTCACCGCTCTTTTAATG 153
QY 692 GCATCGGACACGCTGCGCGTATCGTGGCGCGCATCCGTCGGAAGAGATCAAGTCG 751
Db 154 GTATTGGTGATACGATCCGAATTAGTTTATCAATGATCAATAGCAGAGGTGGAAGTTG 213
QY 752 GTTTCGATATTTGAATCGCTCGGTATCGGTTCGGAGGGATCACTTCATCGCTGCG 811
Db 214 GTAAAGCAATGTAATTCATTAGGACTTTATGATAAATGTTGATATTAATGCTGCC 273
QY 812 CGACCTGTTTCGGCTCAGGAATTTGATTTATCGGTACGTTAAGCGCTGGAGCAACGCC 871
Db 274 CAACATGTGCTGTTAGTAATATGACCTTTTCCCGTGTAAAGAAATTAGAGATATA 333
QY 872 TCGAAGATATCATCACTCCGATGGACGTTTCGATTATCGGCTCGCGTGGTGAATGGCCC 929
Db 334 CAAAAGATTTAAATTTCCATTAAAAATTTGCCATTTTAGGTTGTGTTGTTAATGGTCC 391
RESULT 4
BZ577824
LOCUS 1127 bp DNA linear GSS 17-DRC-2002
DEFINITION msh2 5589.y2 msh Pseudomonas aeruginosa genomic clone msh2_5589,
genomic survey sequence.
ACCESSION BZ577824
VERSION 1 GI:27212885
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1127)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1. .1127
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_5589"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."
BASE COUNT 216 a 381 c 353 g 171 t 6 others
ORIGIN
Query Match 6.7%; Score 74.8; DB 29; Length 1127;
Best Local Similarity 58.5%; Pred. No. 1.8e-11;
Matches 166; Conservative 0; Mismatches 114; Indels 4; Gaps 2;
QY 1 ATGCATAACCAGGCTCCCAATTCACGTAGAAATCAACACGATATTACGTTGGGATGTG 60
Db 712 ATTCATAGCGTTCTCCGATCATTCGCCCAAGTCTCGGAAATCTGGGTCGGCAACGTN 771
QY 61 CCGATTGGCGATGGTCTCCCATCGCGTACAGTCCATACCAATACGCGTAGCAGAC 120
Db 772 CCGGTGGCGCGCATGCGCGATCGCGTGCAGAGATGACCAACACGAGAACTGCGACG 831
QY 121 GTCGAAGCAACGGTCAATCAATCAAGCGCTGGAACGCGTTGGCGCTGATATCGTCCGT 180
Db 832 GTTGTGGCAGCTT--GGCCAAATCCGCGCTCGGAAGATCGCGCGGACATCGTCCG 899
QY 181 GTATCGGTACGACATGAGCGCGGAGAGCGTTCAAACACTCATCAACAGCAGGTTAAC 240
Db 890 GTTTTTCGTCCTCCGACTGGAACCCCGGAAGGGTTTC--GCAAGACCAACCAAGTCAC 947
QY 241 GTGCGCGTGGTGGCTGACATCCACTTCGACTATCGCATTCGCTGCGCT 284
Db 948 CTGCGCTGGTCCCGACATTCATTCATTGACTATTGAATCTCCCT 991

BZ569685 1268 bp DNA linear GSS 17-DRC-2002
pac2-164_970.s1 pac2-164 Pseudomonas aeruginosa genomic clone
pac2-164_970, genomic survey sequence.
ACCESSION BZ569685
VERSION 1 GI:27204646
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 1268)
AUTHORS Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol., (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu

```
source
1. .l323
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="WSH"
/db_xref="taxon:287"
/clone="msh2 4614"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."
240 a 394 c 318 g 371 t

BASE COUNT      240 a 394 c 318 g 371 t
ORIGIN
Query Match      28.5%; Score 319.4; DB 29; Length 1323;
Best Local Similarity 72.0%; Pred. No. 9.7e-87;
Matches 430; Conservative 0; Mismatches 166; Indels 1; Gaps 1;

QY 361 GTGGTTGACTGTGCGCGCATAAAAATTCGGATCCGATTGTCGCTTAACGCCGATCG 420
Db |||||
104 GTGGTCGACGCGCGCGGAGCAACATCCCGATCCGATCCGATCCGATCCGATCCGATCCG 163
QY 421 CTGAAAAGATCTCAAGAAAGTATGGGAACCGACCGCCGAGCGGTGCTGGAATCT 480
Db |||||
164 CTGAAAAGACCTGCGAAGAATAACGGCAACCGACCCCGAAGCCCTGCTCGAATCG 223
QY 481 GCCATCGCTATGTCATCTCGATCGCTGAATTCGATCAGTTCAAAGTCAGCGTG 540
Db |||||
224 GCATCGCGCAGCTGATCATCTCGAAGTGGATTCAGAACTTCAGGTACCGTC 283
QY 541 AAAGCGTCTGACTTCTTCCTCGCTGTGAGTCTTATCGTTTGTGCAAAACAGATCGAT 600
Db |||||
284 AAGGCTCTCGACGCTTTCATGGCGTCTGCGCTATCGCTGTGCGCAGGACAGTCGAG 343
QY 601 CAGCGTTGCATCTGGGATCACCGAAGCGGTGTGCGGACGCGGCGAGTAAATCC 660
Db |||||
344 CAGCCCTCGACCTAGGCATCACCGAGCGCGCGCTGCGCTCCGCGACGTTGAATCG 403
QY 661 GCCATTGTTTAGTCTGTGCTGTCTGAAGGATCGCGACACGCTGCGCGTATCGCTG 720
Db |||||
404 CGGTGGGCTGGGATGCTCTTGGCGGAGGAATCGCGACACCATCCGATTCCTTG 463
QY 721 GCGGCGATCCGGTCGAAGAGATCAAAGTCGGTTTCGATATTTTGAATCGTGCATC 780
Db |||||
464 GCTGCCGATCCGGTCGAGGAGATCAAGTCGGTTTCGACATCTCAAGTCCCTGCACCTG 523
QY 781 CGTCCGAGGATCAACTTCATCGCTGCCCGACCTGTTCGCTCAGAAATTTGATTT 840
Db |||||
524 CGCTCCGTCGATCAACTTCATCGCTGCCCGAGCTGTTTCGCGGAGAACTTCGACGTG 583
QY 841 ATCGGTACGTTAAACGCTGGAGCAACGCTCGAAGATATCATCTCCGATGGACGTT 900
Db |||||
584 GTGAAGACCATGAACAGCTGGAGGGCGCTGGAGACCTGCTGTGGCGATGGACGTG 643
QY 901 TCGATTATCGGCTCGGTGTAATGGCCAGGTGAGCGCTGTTTTCACACTCGCG 957
Db |||||
644 GCCGTAATC-GTTGCGTGAACGTTACGTCGCGGAGCAAGAGGCCATGTTGTC 699

RESULT 2
BU004160/c
LOCUS
DEFINITION
GG37N12.yg.ab1 OG EFHJ lettuce serriola Lactuca sativa cDNA clone
ACCESSION
BU004160
VERSION
BU004160.1 GI:22438555
KEYWORDS
SOURCE
Lactuca sativa
ORGANISM
Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.
REFERENCE
1 (bases 1 to 351)
AUTHORS
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison
, P., Kolkman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z.,
Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenomics.ucdavis.edu/
Unpublished
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]
singleton, see http://cgpdb.ucdavis.edu/ for details.
Plate: QGG37 row: N column: 12.
FEATURES
Location/Qualifiers
1..351
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QGG37N12"
/lab_host="E.coli"
/clone_lib="QG EFHJ lettuce serriola"
/note="Vector: pBRCNDSFIAB; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_SEQ=Not found"
62 a 114 c 115 g 59 t 1 others
BASE COUNT      62 a 114 c 115 g 59 t 1 others
ORIGIN
Query Match      7.3%; Score 81.8; DB 13; Length 351;
Best Local Similarity 53.1%; Pred. No. 6.7e-14;
Matches 173; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 429 AGATCTCGAAGAAAAAGTATGGCAACCGACGCGGCGGTGCTGGAATCTGCCATCG 488
Db |||||
332 AGCTCTCGCCCAAGTAGCGGCGGCCACGCTGAGCGGCTCGTGGCTCGCTCTCAA 273
QY 489 TCATGTTGATCATCTGATCGCTGAATTCGATCAGTTCAAGTCAGCTGAAACGGTC 548
Db |||||
272 GGAAGCGCGCATGTTTCAAGATCTGGGCTTCCCGCTTCAAGATCTCGTCAAGCACA 213
QY 549 TCAGCTCTTCTCGCTGTTGAGTCTTATCGTTTGTGGCAAAACAGATCGATCAGCGTT 608
Db |||||
212 CGGCCCGGTACACATGTTGGAGACCTACCGATCTGCTCGCTCAAGAGGAGACTGTCCT 153
QY 609 GCATCTGGGGATCACCGAAGCGGTGGTGGCGGAGCGGGGAGTAAATCCGCCATTGG 668
Db |||||
152 GCACCTGGAGTCACTGAGCAGGTCGCGCTGGCAGGCCCGCCATCAAGTCTTGGCTGC 93
QY 669 TTTAGTCTGCTGCTGTTGAAGGATCGGCACACGCTGCGCGTATCGCTGGCGGCCGA 728
Db |||||
92 ATTGCTGCGCTGCTGCGCGAAGGATCGGCACACGATTGGGGTCTTCTCTCNCGCC 33
QY 729 TCCGTCGAAGAGATCAAAAGTCGGTT 754
Db |||||
32 TCTGTGGAGAGGTGAAGTGGGT 7

RESULT 3
BH235010
LOCUS
DEFINITION
PH_05.x Spiroplasma kunkelii H Spiroplasma kunkelii genomic clone
ACCESSION
BH235010
VERSION
BH235010.1 GI:18030478
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 13:42:06 ; Search time 2752 Seconds
(without alignments)

9882.525 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119

Sequence: 1 atgcataaccaggctccaat.....ttcagcaggttgaaaaataa 1119

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estnu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_est3:*

12: gb_est4:*

13: gb_est5:*

14: gb_est6:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_tod:*

26: em_gss_phg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	319.4	28.5	1323	29	BZ575759 msh2_4614
2	81.8	7.3	351	13	BU004160 QGG37N12
3	81.6	7.3	720	28	BH235010 PH_05.x S
4	74.8	6.7	1127	29	BZ577824 msh2_5589

C	5	68.2	6.1	1268	29	BZ569685
	6	67.8	6.1	448	9	AV644003
	7	67.4	6.0	482	9	AV642898
	8	67.4	6.0	502	9	AV626792
	9	67.4	6.0	512	9	AV644476
	10	67.4	6.0	531	9	AV628026
	11	67.4	6.0	531	9	AV644087
	12	67.4	6.0	545	9	AV626844
	13	65.8	5.9	467	9	AV629427
	14	62	5.5	377	9	AV641168
	15	58	5.2	688	14	CD424926
	16	58	5.2	760	14	CD423530
	17	56.8	5.1	792	14	CB620208
	18	56.8	5.1	821	14	CB628414
	19	56.8	5.1	852	14	CB628479
	20	56.6	5.1	658	14	CD233378
	21	56.4	5.0	380	9	AI770308
	22	54.2	4.8	303	9	AV640429
	23	53.8	4.8	597	9	AW282425
	24	53.6	4.8	728	10	BG525868
C	25	53.4	4.8	658	28	AZ570993
	26	52.6	4.7	640	9	AW738440
	27	52.6	4.7	680	12	BG890681
	28	51.8	4.6	731	10	BG124857
	29	51.6	4.6	748	14	CB662233
	30	51.6	4.6	770	14	CB662235
	31	51.6	4.6	828	14	CB669733
	32	51.6	4.6	853	14	CB619369
	33	51.4	4.6	492	12	BM323500
	34	51.4	4.6	580	14	CD423350
	35	51.4	4.6	722	14	CA483211
	36	51	4.6	579	12	BI935552
	37	51	4.6	700	12	BI933447
	38	50.4	4.5	173	10	BF962770
	39	50	4.5	662	14	CB662234
	40	49.8	4.5	576	13	BQ975044
	41	49.6	4.4	489	6	AUI86794
	42	49.4	4.4	532	10	BE924238
	43	49.4	4.4	645	12	BI934498
	44	48.4	4.3	558	9	AV917069
	45	48.4	4.3	603	12	BJ551332

ALIGNMENTS

RESULT 1

BZ575759

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ575759 1323 bp DNA linear GSS 17-DEC-2002
msh2_4614.x1 msh Pseudomonas aeruginosa genomic clone msh2_4614,
genomic survey sequence.

BZ575759 1 GI:27210820

GSS.

Pseudomonas aeruginosa

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 1323)

Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,

Burns,J.L., Kaul,R. and Olsen,M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol., (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

FEATURES

THIS PAGE BLANK (USPTO)

Job time : 372 secs

QY 918 GGTGAATGGCCAGGTGAGCGCTGCTTTCTACACTCGGCGTCACCGGGGCAACAAGAA 977
 DB 912 GGTACAGGCCCTGGAGAGCTGCGAAGCGGATATCGGATGCTGGTTCACACGGAGA 971
 QY 978 AGCGGCCCTCTATGAAGATGG 998
 DB 972 AGGCCTTCTTTTATGACATGG 992

RESULT 15

ABA03041

ID ABA03041 standard; DNA; 2944528 BP.

XX AC ABA03041;

XX DT 05-FEB-2002 (first entry)

XX DE Listeria monocytogenes EGD-e genome sequence.

XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;

XX KW vitamin B12; bacterial infection; disease; ds.

XX OS Listeria monocytogenes.

XX PN WO200177335-A2.

XX PD 18-OCT-2001.

XX PF 11-APR-2001; 2001WO-FR01118.

XX PR 11-APR-2000; 2000FR-0004629.

XX PA (INSP) INST PASTEUR.

XX PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Feihi H, Dehoux P;

XX PI Dussurget O, Chetoui F, Nedjari H, Glaser P, Kunst F, Cossart P;

XX PI Daniels J, Goebel W, Krefit J, Kuhn M, Ng E, Vazquez-Boland JA;

XX PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;

XX PI Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;

XX PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;

XX PI Maduenio E, De Pablo B, Wehland J, Kaerst U, Entian K, Hauf J;

XX PI Rose M, Voss H;

XX WPI; 2002-010914/01.

XX PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment

XX PT and prevention of Listeria and related bacterial infections, and

XX PT related polypeptides

XX PS Claim 1; SEQ ID No 1; 192pp; French.

XX CC The present sequence is the genome sequence of Listeria monocytogenes

XX CC EGD-e. This sequence and fragments of this sequence are useful for

XX CC selecting probes and primers for detecting genes in L. monocytogenes and

XX CC related organisms, and to study genetic polymorphisms and other genomes.

XX CC Proteins (ABR47297-ABR50149) expressed from the present sequence are

XX CC useful for raising specific antibodies, identification of L.

XX CC monocytogenes and related organisms, and for biosynthesis and

XX CC biodegradation, especially biosynthesis of Vitamin B12. This sequence and

XX CC proteins encoded by it are also useful for selecting compounds that

XX CC regulate gene expression and cell replication and modulate L.

XX CC monocytogenes-related diseases. In addition, this sequence and proteins

XX CC encoded by it are useful in pharmaceutical and vaccine compositions for

XX CC the treatment or prevention of infections by L. monocytogenes and related

XX CC organisms.

XX CC Note: The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 2944528 BP; 914202 A; 563301 C; 555061 G; 911964 T; 0 other;

Query Match

22.5%; Score 251.4; DB 24; Length 2944528;

Search completed: November 23, 2003, 13:55:27

Best Local Similarity 53.5%; Pred. No. 2.8e-61;
 Matches 525; Conservative 0; Mismatches 456; Indels 0; Gaps 0;

QY 18 AATTCAAGTGAATAATCAACACAGTATTACGTTGGGATGTCCGATTTGGCGATGGTGC 77
 DB 1474813 AATATTTCCGGAACACATCGCCAGTCCAGTCAATTTAACTATTGGTGGTAGTGA 1474872

QY 78 TCCCATCGCGGTACAGTCCATGACCAATACGCGTACGACAGACAGTCCGAAACCGTCAA 137
 DB 1474873 GGAATTAATCAATCCAAAGCATGACTACTACAAGACACATGATGTCGAAGCAACAGTAGC 1474932

QY 138 TCAATCAAGGCGCTGGAAACGCGTTGGCGCTGATATCGTCCGTATCGTATCCGATCCGAT 197
 DB 1474933 AGAAATTCACCGATTAGAAAGAGTGGTTGTGAGATTTGCGAGTTGCTTGTCTCGATGA 1474992

QY 198 GGACGGCGCAGAGCGTTCCAACTCATCAAAACAGCAGGTTAAACGTCGCCCTGGTGGCTGA 257
 DB 1474993 ACGTGACGGAATGCTTTAGTGCCATCAAGAAAAGATTCATATTCGCCCTTTGGTGGAGA 1475052

QY 258 CATCCACTTCGACTATCGCATTCGCGTCAAAAGTAGCGGAATACGGCGTTCGATTTGTCTGCG 317
 DB 1475053 TATTCATTTGATTACCGACTAGCACTTAAAGCTATTGATGACGAGTGTGACAAATTCG 1475112

QY 318 TATTAACCTCGCAATATCGGTAATGAAGAGCGTATTCGCATGCTGGTGTGATCTGTCGCGG 377
 DB 1475113 GATTAATCCTGGTAACATTTGGTCCGCGTATCGGGTGGAAAAAGTGTAAATGCTCTAA 1475172

QY 378 CGATTAACATTCGATCCGTTATTTGGGTTTAAACCGGATCGCTGGAAGAAAGATCTGCA 437
 DB 1475173 AGCAAAAAATATTTCCAAATCCGTTATTTGGGTTTAAATGCTGGTAGTTTGAAGAAAGAAATAT 1475232

QY 438 AGAAAGATGATGGCAACCGACCGCGTTGCTGGAATCTGCGCATCGCTCATGCTTCA 497
 DB 1475233 TCAGAAATATGTTTACCCTACTGCTGACGGAATGTTAGAAAGTGCATTTGCCATATTA 1475292

QY 498 TCATCTCGATCCGCTGAATTCGATTCAGTTCAAAGTACAGCGTGAAGCGTCTGACGCTTT 557
 DB 1475293 AATTCTCGAAGATTTAGATTTTATGATATCATTTGTTCTTTTGAAGGCTTCTGATGTGA 1475352

QY 558 CTTGCTGTTGAGTCTTATCTGTTGCTGGCAAAACAGATCGATCAGCGCTTCGATCTGGG 617
 DB 1475353 TTTAGCAATTTGAAGCTTATGATAAAGCTAGTCGCGCATTTAAATTTATCTCTGATCTCGG 1475412

QY 618 GATCACCAAGCGCGTGGTGGCGGCGGCGGAGTAAATCCGCCATTTGGTTTAGGTCT 677
 DB 1475413 AATTACAGATCTGGTACACAAATTTGCTGGAGGAATAAAGTGTCTGTTTAGAGC 1475472

QY 678 GTGCTGTCTGAAGGATCGGCGACACGCTCGCGTATCGTGGCGGCGGATCCGGTTCGA 737
 DB 1475473 GATACCTAGTTTGGGCAATTTGGAATACATTAACGAGTATCTTTGAGTGTCTGATCTGGA 1475532

QY 738 AGAGATCAAGTCGGTTTCGATATTTTGAATCGTCCGTATCCGTTCCGAGGGATCAA 797
 DB 1475533 AGAAATAAAGTGGCGCGGGAAGTTTAAATTCATTTGTTGCTTTCTCGAATGCCGCAAT 1475592

QY 798 CTTATCGCTCCCGACCTGTTTCGCGTCAAGAAATTTGATTTATCGGTACCGTTAAACGC 857
 DB 1475593 GCTTATCTCTGCGCTACTTGGCGTCAATAGATTTGATTTAAATTCGTATCGTAAATGA 1475652

QY 858 GCTGGAGCAACCGCTGGAGATATCATCTCCGATGGAAGTTCGATTTATCGGCTGCGT 917
 DB 1475653 AGTGGAAATTTACATAGCAAGATTTGAAGTTCGATTTAAAGTAGCGGTGCTTTGGCTGTGC 1475712

QY 918 GGTGATGCCAGGTGAGCGCTGGTTTCTACACTCGCGCTCACCGGCGGCAACAAGAA 977
 DB 1475713 GGTCAACGCGCTCGAGAGCTCGCAAGCCGATATCGGAATTTGCTTCAACACGAGA 1475772

QY 978 AAGCGGCTCTATGAAGATGG 998
 DB 1475773 AGGCCTTCTTTTATGACATGG 1475793

Db 552 TTTAGCAATTGAAGCTTATGATAAAGCTAGTCGCGCATTTAATATCTCTGCATCTCGG 611
 Qy 618 GATCACCAGACGGGTGGCGCGAGCGGCGAGTAAATCCGCATTCGTTAGGTCT 677
 Db 612 AATTACAGAACTGGTACAAATTTGCTGGAGGAATAAAAGTCTGCTGTTTAGGAGC 671
 Qy 678 GCTGCTGTCTGAAGGCGATCGGCACACGCTGCGCGTATCGCTGGCGCGGATCCGCTCGA 737
 Db 672 GATACTCAGTTTGGGCATTGGAAATACATTTACGAGTATCTTTGAGTGTCTGTGA 731
 Qy 738 AGAGATCAAGTCGGTTTCGATATTTTGAATTCGTCGCTGATCCGTTGCGAGGATCAA 797
 Db 732 AGAAATAAAGTGCGCGGGAAGTTTAAATCATTTGGTCTTTCCCTCGAATGCCCAAT 791
 Qy 798 CTTTCATCGCTGCCGACCTGTTGCGGTGAGGATTTGATGTTATCGGTACGGTTAACGC 857
 Db 792 GCTTATCTCTGCGCTTACTTGGCGTGAATAGAGATTTGATTTAATTCGTATCGCTAATGA 851
 Qy 858 GCTGGAGCAACGCTCGAAGATATCATCTCCGATGGAGCGTTTCGATTATCGGCTGCGT 917
 Db 852 AGTGGAAATACATAGCAAGATTGAAGTTCCGATTAAGTAGCGTGTGGCTGTGC 911
 Qy 918 GGTGAATGCCAGGTCGAGCGCTGTTTCTACACTCGCGGTACCGCGGCGCAACAGAA 977
 Db 912 GGTCAACGGCCCTGGAGAAGCTCGGAAGCCGATATCGGAATTTGCTGTTCAACCGGAGA 971
 Qy 978 AAGCGCCCTTATGAAGATGG 998
 Db 972 AGGCCTTCTTTTAGACATGG 992

RESULT 14

ABQ69993
 ID ABQ69993 standard; DNA; 1107 BP.
 AC ABQ69993;
 XX
 XX 29-AUG-2002 (first entry)
 XX
 XX Listeria monocytogenes EGDe DNA sequence #205.
 DE
 XX
 KW Antibacterial; Listeria; food contamination; mutational analysis;
 KW infection; ds.
 XX
 OS Listeria monocytogenes EGDe.
 XX
 PN WO200228891-A2.
 XX
 PD 11-APR-2002.
 PF 04-OCT-2001; 2001WO-FR03061.
 PR 04-OCT-2000; 2000FR-0012697.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Kunet F, Glaser P;
 XX
 DR WPI; 2002-332479/37.
 XX
 PT New genomic sequences from Listeria species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators -
 XX
 PS Claim 16; SEQ ID 2806; 180pp; French.
 XX
 CC The present invention relates to nucleic acid sequences
 CC (ABQ67188-ABQ71212) from Listeria sp. The sequences are useful as probes
 CC and primers for identification and/or detection of Listeria (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be

CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of Listeria (potential therapeutic agents), also for
 CC treating infections by Listeria, and are useful as immunogens in
 CC anti-Listeria vaccines.
 CC Note: the sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1107 BP; 348 A; 178 C; 255 G; 326 T; 0 other;
 Query Match 22.5%; Score 251.4; DB 24; Length 1107;
 Best Local Similarity 53.5%; Pred. No. 9.7e-63;
 Matches 525; Conservative 0; Mismatches 456; Indels 0; Gaps 0;
 Qy 18 AATTCAACGTAGAAATCAACACAGTATTACGTTTCGGGAATGCGCGATTGGCGATGGTGC 77
 Db 12 AATATTTTCGGAACACACTCGCCAGTCCAGTGGGTAAATTTAACTATTGGTGGTAGTGA 71
 Qy 78 TCCCATCCCGGTACAGTCCATGACCAATACGCTACGACAGACGTCGAAGCAACGCTCAA 137
 Db 72 GGAATTAATCTCCAAAGCATGACTACTACAAGACACATGATGTCGAAGCAACAGTAGC 131
 Qy 138 TCAAAATCAAGCGCTGGAAACGCGTTGGCGCTGATATCGTCCGTGTATCCGTACCGACGAT 197
 Db 132 AGAAATTCACCGATTAGAAGAAGCTGGTTGTTCAGATTGTCGAGTTGCTTGTCTGATGA 191
 Qy 198 GGACCGGCGAGAGCGTTCAAACTCATCAACAGCAGGTTAACGTGCCGCTGGTGGCTGA 257
 Db 192 AGTGCAGCGAATGCTCTTAGTGCCATCAAGAAAAGATTATATTCGCTTGTGGCAGA 251
 Qy 258 CATCCACTTCGACTATCGCATTCGCTGAAAGTAGCGAATACCGCGTCGATTGCTCGG 317
 Db 252 TATTCATTTTGTATACCGACTAGCACTTAAGCTATTGATGCGAGTGTGACAAATTCG 311
 Qy 318 TATTAAACCTGGCAATATCGGTAATGAAGAGCGTATTCGCATGTTGGTGTGATGTCGCGG 377
 Db 312 GATTAATCTGTTAACTTTGGTCCGCTGATCGGGTGGAAAAGTGGTTAATGCTGCTAA 371
 Qy 378 CGATAAAACATTCGATCCGTTATGGCGTTAAACCGCGATCGCTGGAAAAGATCTGCA 437
 Db 372 AGCAAAAATATTCCAATCCGTTATTTGGGGTTAACTCTGTTAGTTTGAAGAAAATTA 431
 Qy 438 AGAAAGTATGCGAACCGACCGCGCGCGCTGCTGGAATCTGCGCATGCGCTCATGTTGA 497
 Db 432 TCAAAAATATGTTTACCCTACTCTGACGGAATGTTAGAAAGTGCATCTGCCCATTA 491
 Qy 498 TCATCTCGATCCCTGAACTTCGATCAGTTCAAAAGTCAGCGTGAAGCGTCTGACGCTTT 557
 Db 492 AATTCTCGAAGATTAGATTTTATGATATCATTTGTTCTTTGAAGGCTTCTGATGTGA 551
 Qy 558 CTGCTGTTGAGTCTTATCGTTTCTGGGCAACACAGATCGATCAGCGTTGCATCTGGG 617
 Db 552 TTTAGCAATTGAAGCTTATGATAAAGCTAGTCGCGCATTTAATATCTCTGCTGCTCGG 611
 Qy 618 GATCACCAGACGGGTGGCGCGAGCGGCGAGTAAATCCGCATTCGTTTAGGTCT 677
 Db 612 AATTACAGAACTGGTACAAATTTGCTGGAGGAATAAAAGTGTCTGCTGTTTAGGAGC 671
 Qy 678 GCTGCTGTCTGAAGGCGATCGGCGACACGCTGCGGTATCGCTGGCGCGGATCCGCTCGA 737
 Db 672 GATACTCAGTTTGGGCATTGGAAATACATTTACGAGTATCTTTGAGTGTCTGATCTCTG 731
 Qy 738 AGAGATCAAGTCGGTTTCGATATTTTGAATTCGTCGCTGATCCGTTGCGAGGATCAA 797
 Db 732 AGAAATAAAGTGCGCGGGAAGTTTAAATCATTTGGTCTTTCCCTCGAATGCCCAAT 791
 Qy 798 CTTTCATCGCTGCCGACCTGTTGCGGTACAGAAATTCATGTTATCGGTACGGTTAAACGC 857
 Db 792 GCTTATCTCTGCGCTTACTTGGCGTGAATAGAGATTTGATTTAATTCGTATCGCTAATGA 851
 Qy 858 GCTGGAGCAACGCTGGGAAGATATCATCTCCGATGGAGCGTTTCGATTATCGGCTGCGT 917
 Db 852 AGTGGAAATTTACATAGCAAGATTGAAGTTCCGATTAAGTAGCGTGTGGCTGTGC 911

(I) is useful in DNA arrays or chips to carry out analysis of the expression of the Bifidobacterium gene. ABQ81844 to ABQ81850 represent Bifidobacterium related nucleotide sequences given in the Sequence Listing from the present invention but not mentioned further within the specification.

N.B. The sequence data for this patent is not represented in the printed specification but is based on sequence information supplied by the European Patent Office.

Sequence 349980 BP; 72540 A; 102738 C; 103221 G; 71481 T; 0 other;

Query Match 27.4%; Score 306.2; DB 24; Length 349980;
Best Local Similarity 55.9%; Pred. No. 1.3e-77;
Matches 603; Conservative 0; Mismatches 473; Indels 3; Gaps 1;

QY 25 CGTAGAAATCACAGGTATTTACGTTGGATGTCGCGATGGCGATGGCTCCCATC 84
DB 119782 CGCCGAAGTCCCGCGCATATATGGTGGGTGCGGTGCCGTGGGGGAGCGCCCATC 118841
QY 85 GCGGTACAGTCATGACCAATACGCGTACGACAGAGTCGAAACACGCGTCAATCAAATC 144
DB 118842 TCGGTGCAATGATGACCAACAGCTCAGGCCAAGTACCTGCCAGCTTGACGAGATC 118901
QY 145 AAGGCGCTGGAACGCGTGGCGTGTATGTCGTCGTCATCCGACGATGAGACGG 204
DB 118902 GCGAACTGACCGCGCGCTGCGACATGTCGCGTGGCGTCCGAGTCAGGACGAT 118961
QY 205 GCAGAGCGTTCAAACTCATCAACAGCAGGTAAAGTCGCGTGGTGTGACATCCAC 264
DB 118962 GCGAGCGCACTGCCGGAATCTGCCGAAATCCGCGATTCGCGTATCCGATATCCAC 119021
QY 265 TTCGATATCGCATTCGCTGAAAGTAGCGGATAGCGGTTCGATTGTCTGCGTATTAAC 324
DB 119022 TTCAGTCAAATACGTGTTCAGGCCATCAGCGCGGTGCGCGCGTGGCGTGAAC 119081
QY 325 CTGCGAATATCGGTAAATGAAGCGGTATTCGATGG---TGTTGACTGTGCGCGCGAT 381
DB 119082 CCGGCAACATCGTAAATGTCGACGAAGTCGCGCGGACATCTGAAGCGCGCCACCGAC 119141
QY 382 AAAAAATTCGATCGTATGCGTTTACGCGGATCGCTGAAAGATCTGCAAGAA 441
DB 119142 GCAGGCACTCGCTGCGTATCGGTGTGAACCGCGCTGCGTGAAGAACTCTACGC 119201
QY 442 AAGTATGGCAACCGACGCGCGAGGCTTCTGGAATCTGCCATCGCTCATGTGATCAT 501
DB 119202 AATACGTTGCCCGACTCCGAGAGCGTGTGGCTCCGCTTGAAGAGCGACATG 119261
QY 502 CTCGATCGCGTGAATTCGATCAGTTCAAAGTCAGCGTGAAGCGTCTGACGTCTTCCTC 561
DB 119262 TTCGAGGACGTGGCTTCCAGCATTTCAAGATCTCGTCAACACCATGACGTCAACCC 119321
QY 562 GCTGTTGAGTCTTATCGTTTGTGCGCAACACAGATCGATCAGCGTTCGATCTGGGATC 621
DB 119322 ATGTTGCAAAACGTACCGACTCTCGTTTCAAGGGGATTTGGCGGTGCACTTCGGGTC 119381
QY 622 ACCGAGCGCGTGGTGGCGCGAGCGGGAGTAAATCCGCCATTTGTTAGTCTGCTG 681
DB 119382 ACCGAGCGCGCTGTTGGCAAGCACCATCAATCTGCTGCGTTCGTTGCGCTC 119441
QY 682 CTGTTGAAGGATTCGCGACAGCTGCGGTATCGCTGCGGCGCGATCCGTTGGAAGAG 741
DB 119442 TTGGCGAGGCAATGGCGACAGATTCGCGGTGTCTCTCTCCGCGCGCGCGCGAAGAG 119501
QY 742 ATCAAGTCGTTTCGATTTTGAATTCGCTGATTCGTTGCGGAGGATCACTTC 801
DB 119502 GTCAAGTGGTTTGCAAACTCTTGAATATCATGGATTGCGCGCCCGCAAGTTCGATC 119561
QY 802 ATGCGCTGCCGACCTGTTCCGCGTCAAGAAATTCGATTCGATTCGATTCGATTCGATTC 861
DB 119562 ATCTCTGCGCGAGCTGCGCGCGCGCCCAAGTGAATGTATCAGCTCGCTCGCGCTC 119621
QY 862 GAGCAACGCTGGAAGATATCATCACTCCGATGCAAGCTTTGATTTATCGCTCGTGGTG 921

DB 119622 ACCGAAGCCTTGAAGGACGTGACCGCGCGGATCCGAGTGGCGGTTCATGGCTGCATCGTC 119681
QY 922 AATGCCCCAGGTGAGCGCGCTGTTTCTACTCGGCGGTCAACGGCGCAACAAGAAAGC 981
DB 119682 AACGTCGCGCGAGGACGCGAAGCCGACCTCGGCTGGCTCGGCAATGGCAAGGC 119741
QY 982 GGCCTCTATGAATGCGGTGCGGAAAGACCGTCTGGACACAACAGATATGATCGACCG 1041
DB 119742 CAGATCTTTCATCAAGGCGAAGGTTCATCAAGACCGTGCCTGCGAAGACGATCGTCGACAG 119801
QY 1042 CTGGAAGCACGATTCGTCGGAAGCCAGTCAGCTGAGGACGCGCTCGAATTGACGT 1100
DB 119802 CTGTCACCATAGCAACGATATCGCGCCCAATGGAGCGCGACGTTCAAGTCCCCGT 119860
RESULT 11
ABK74401
ID ABK74401 standard; DNA; 1083 BP.
XX
AC ABK74401;
DT 13-AUG-2002 (first entry)
XX
DE Bacillus licheniformis genomic sequence tag (GST) #1692.
XX
XX Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.
XX
OS Bacillus licheniformis.
XX
PN WO200229113-A2.
XX
PD 11-APR-2002.
XX
PF 05-OCT-2001; 2001WO-US31437.
XX
PR 06-OCT-2000; 2000US-0680598.
PR 27-MAR-2001; 2001US-279526P.
XX
XX (NOVO) NOVOZYMES BIOTECH INC.
PA (NOVO) NOVOZYMES AS.
XX
PI Berka R, Clausen IG;
XX
XX WPI; 2002-416684/44.
DR
XX
PT Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array
XX
PS Claim 4; SEQ ID NO 1692; 200pp; English.
XX
CC The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring of the way
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions.
CC environmental stress or other physiological provocation. Extensive
CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.

XX DNA encoding novel human diagnostic protein #25870.
DE Human; Chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX Homo sapiens.
XX WO200175067-A2.
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI: 2001-639362/73.
XX P-PSDB; ABG25879.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 1; SEQ ID No 25870; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human
CC diagnostic coding sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1857 BP; 463 A; 472 C; 530 G; 392 T; 0 other;
Query Match 31.8%; Score 356; DB 23; Length 1857;
Best Local Similarity 97.9%; Pred. No. 4.4e-93;
Matches 466; Conservative 0; Mismatches 0; Indels 10; Gaps 10;
QY 644 GCGGGGCGAGTAAATCCCGCATTTAGTGTCTGCTGTCTGAAGGCATCGGCGACA 703
DB 512 GCGGGGCGAGTAAATCCCGCATTTAGTGTCTGCTGTCTGAAGGCATCGGCGACA 570
QY 704 CGCTGGCGGTATCGTGGGCGGCGATCCGGTCCGAAGAGATCAAGTCGGTTTCGATATT 763
DB 571 CGCTGGCGGTATCGTGGGCGG-CGATCCGGTCCGAAGAGATCAAGTCGGTTTCGATATT 629
QY 764 TGAATTCGCTGATCCGTTCCGAGGGATCAACTTCATCCGCTCCGACCTGTTCCG 823
DB 630 TGAATTCGCTG-GTATCCGTTCCGAGGGATCAACTTCATCCGCTCCGACCTGTTCCG 688
QY 824 GTCCAGGAATTTGATGTTATCGGTACGGTTAAGCGCTGGAGCAAGCCCTGGAGGATATCA 883

DB 689 G-CAGGAATTTGATGTTATCGGTACGGTTAAGCGCTGGAGCAACGCCCTGG-AGATATCA 746
QY 884 TCACCTCCGATGACGTTTCGATTATCGGCTGGTGAATGCGCCAGGTGAGGCGCTGG 943
DB 747 TCACCTCCGATGACGTTTCGATTATCGGCTGGTGAAT-GCCAGGTGAGGCGCTGG 805
QY 944 TTTCTACACTCGGCTCACCGCGGCAACAAGAAAGCGGCTCTATGAAGATGGCGTGC 1003
DB 806 TTTCTACACTCGGCTCACCGCGGCAACA-AAAAGCGGCTCTATGAAGATGG-GTGC 863
QY 1004 GCAAGACCGTCTGGACAAACAGATATGATCGACAGCTGGAAGCACCATTCTGTCGA 1063
DB 864 GCAAGACCGTCTGGACAAACA-GATATGATCGACAGCTGGAAGCACCATTCTGTCGA 922
QY 1064 AAGCCAGTCAGCTGCAAGGCGGCTCGAATTTGACGTTTCAGCAGGTTGAAAAATAA 1119
DB 923 AAGCCAGTCAGC-GGACGAAGCGGCTCGAATTTGACGTTTCAGCAGGTTGAAAAATAA 977
RESULT 10
ABQ81842
ID ABQ81842 standard; DNA; 349980 BP.
XX
AC ABQ81842;
XX
DT 19-NOV-2002 (first entry)
XX
DE Bifidobacterium longum NCC2705 genomic sequence SEQ ID NO:1.
XX
KW Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
KW antidiarrheic; antibacterial; inhibitor of Salmonella; detection;
KW identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
KW rotavirus; food composition; pharmaceutical composition; gene; ds.
XX
OS Bifidobacterium longum.
XX
PN EP1227152-A1.
XX
PD 31-JUL-2002.
XX
XX 30-JAN-2001; 2001EP-0102050.
XX
XX 30-JAN-2001; 2001EP-0102050.
XX (NEST) SOC PROD NESTLE SA.
XX
XX WPI; 2002-668397/72.
XX
PT Novel polynucleotide comprising Bifidobacterium genome sequence useful
PT as a probe or primer for detecting and/or identifying Bifidobacterium
PT longum in a biological sample -
XX
PS Claim 1; SEQ ID 1; 80pp; English.
XX
CC The present invention describes a polynucleotide (I) comprising a
CC sequence of a Bifidobacterium genome selected from the nucleotide
CC sequences given in ABQ81842 and ABQ81843, or a sequence exhibiting at
CC least 90% identity or which hybridises with the sequences given in
CC ABQ81842 and ABQ81843. Also described is a polynucleotide (II) encoding
CC a fusion protein, comprising a sequence selected from 1097 sequences
CC given in ABP65258 to ABP66354 ligated in frame to a polynucleotide
CC encoding a heterologous polypeptide. (I) has antidiarrheic and
CC antibacterial activities, and can be used as an inhibitor of Salmonella.
CC (I) (which is a probe) is useful for the detection and/or identification
CC of Bifidobacterium longum in a biological sample. A carrier containing
CC the lactic acid bacterium Bifidobacterium longum NCC2705 (NCIM 1-2618)
CC can be used for preventing and/or treating diarrhoea brought about by
CC pathogenic bacteria and/or rotavirus. The carrier is a food composition
CC selected from milk, yogurt, curd, cheese, fermented milks, milk based
CC fermented products, ice-creams, fermented cereal based products, milk
CC based powders, infant formula, pet food or a pharmaceutical composition
CC selected from tablets, liquid bacterial suspensions, dried oral
CC supplement, wet oral supplement, dry tube feeding or wet tube feeding.

QY 979 AGCGGCTCTATGAAGATGGCGT---GCGCAAGACCGCTCTGGACAACACGATGATC 1035
|||
Db 315253 ATGCAATTTATGACGCGAGTAAGACAAAGAAATAAAACAGAAATTATA 315312
|||
QY 1036 GACCACTGGAAGCACGCAATCGTGGAAA 1065
|||
Db 315313 GAAAAATGGAATTAATAATTCGAAAAAA 315342
|||
RESULT 8
AAF28536
ID AAF28536 standard; DNA; 33140 BP.
XX AAF28536;
AC
XX 04-APR-2001 (first entry)
DT
XX Genomic fragment #23.
DE
XX Genomic library; bacteria; human upper airway; otitis media; sinusitis;
KW bronchopulmonary; endocarditis; meningitis; ss.
KW
XX Moraxella catarrhalis.
OS
XX WO200078968-A2.
PN
XX 28-DEC-2000.
PD
XX 16-JUN-2000; 2000WO-US16649.
PF
XX 18-JUN-1999; 99US-0140121.
PR
XX (INCY-) INCYTE GENOMICS INC.
PA
XX Lagace RE, Patterson C, Berg KL;
PI WPI; 2001-041427/05.
DR
XX Genomic library for identifying diagnostic and therapeutic
PT compositions, and for identifying virulence factors, regulatory
PT elements and drug targets, comprises Moraxella catarrhalis nucleic
PT acids -
XX Claim 1; Page 191-199; 545pp; English.
XX The present invention relates to a Moraxella catarrhalis genomic library
CC comprising of a combination of 41 nucleic acid molecules (see
CC AAF28514-AAF28554). The library has a number of uses described in the
CC specification e.g. is useful for identifying diagnostic and therapeutic
CC compositions. M. catarrhalis (Branhamella catarrhalis) is a large
CC aerobic, gram-negative diplococcus, normally found among the bacterial
CC flora of human upper airways. M. catarrhalis is known to cause acute,
CC localised infections such as otitis media, sinusitis and bronchopulmonary
CC infection and life-threatening, systemic diseases including endocarditis
CC and meningitis.
XX Sequence 33140 BP; 9472 A; 6496 C; 7384 G; 9787 T; 1 other;
SQ
Query Match 34.2%; Score 382.6; DB 22; Length 33140;
Best Local Similarity 59.2%; Pred. No. 2.7e-100;
Matches 652; Conservative 0; Mismatches 449; Indels 0; Gaps 0;
QY 1 ATGCATAACGAGCTCAATTAACGTAGAAAATCAACAGTATTACGTTGGGAATGTG 60
|||
Db 21735 ATGCAATACAAATCCGATTAACGCGCCCTACCAAAAAATCTATGTTGGCAATGTC 21794
|||
QY 61 CGGATTGGGATGCTCCATCGCGTACAGTCCATGACCAATACGCGTAGACAGAC 120
|||
Db 21795 GCATCGGCGGCGATGACCCATCAGCGTCAAAAGTATGCAATACAAACACTGTGAT 21854
|||
QY 121 GTCGAAGCAACGGTCAATCAAAATCAAGGCGCTGGAACGGCGTGGCGGTGATCGTCCGT 180
|||
Db 21855 ATTGATGCAACAGTTGCACAAATTCAGCGATGCGTGGACCGAGTGGCGACTTGGTGGT 21914
|||

QY 181 GTATCGGTACCGACGATGGACGCGGAGAGCGTTCAAACCTCATCAAAACAGCAGGTAAAC 240
|||
Db 21915 GTATCGACACCAACGATGGAGAGCGTGCCTGCTTTGTGAGATAAAAAACGCGTATCC 21974
|||
QY 241 GTGCCGCTGGTGGTGAACATCCACATTCGACTATCGCATTTGCGCTGAAAGTACGGGAATAC 300
|||
Db 21975 ATTCCACTCATTTGCCGATATTCAATTTGATCAAAAAATCGCCATTCGAGTGGCGGATGTG 22034
|||
QY 301 GCGGTGATTTGCTCGGTATTAAACCTCGCAATATCGGTAATGAAGAGCGTATTCGCATG 360
|||
Db 22035 GGTGACAGCTGCTCGGTATCAATCAGGTAATATTGGCAATGATCAAAAAGTCAAAAGAA 22094
|||
QY 361 GTGTTGACTGTGCGCGGATAAAAACATTCGATCCGTTATTTGGCGTTAAACGCGGATCG 420
|||
Db 22095 GTGTTGACGCGGACGCCATCATTAATGTGCCGATTGTTATCGTGTGATGCGAGCTCA 22154
|||
QY 421 CTGAAAAAGATCTGCAAGAAAAAGTATGGCGAACCGACCGCGGCGGTTGCTGGAATCT 480
|||
Db 22155 CTTGAAAAAGGATTTACAAAAAAATATCGAGAACCGACAGGTGAGCGGATGCTAGAATCG 22214
|||
QY 481 GCGATGCGTCAATGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCACGCTG 540
|||
Db 22215 GCGTTACGCCACATTGATATTTTAGAAAAATCTTAATTTTCAGAAATATAAAATCTCAGTC 22274
|||
QY 541 AAAGCGTCTGACGCTCTTCTCGCTGTTGAGTCTTATCGTTTGTGTCGCAAAAACAGATCGAT 600
|||
Db 22275 AAAGCTCAATGATTTTAAACCTTGTATCGTATCGATGATTTTCAGCACAAAATGAT 22334
|||
QY 601 CAGCGCTTGATCTCGGGATCACCGAAGCGGTGTCGCGGAGCGGCGGAGTAAATCC 660
|||
Db 22335 AATCCATTGATCTTTGGCGTTACCGAAGCGGCGTATATCGCACCTGGCAGCGTCAAAATCT 22394
|||
QY 661 GCGATTGTTTAGTCTGCTGCTGTTGAGGCAATCGGACACGCTGCGCTATCGCTG 720
|||
Db 22395 GCGATTGCTTTGGGCGGTTTGTCTTGTGATGCAATGGCGGATACCAATCGTATCTCTTTG 22454
|||
QY 721 GCGGCGGATCCCGTCAAGAGATCAAAGTCGGTTTCGATATTTTGAATTCGCTGCGTATC 780
|||
Db 22455 GCGGACGACCAAGAGAGATTAATTTGCTTTGACATTTAAATCACTGGGTATT 22514
|||
QY 781 CGTTCGCGAGGATCAACTTCATCGCTGCCGACCTGTTCCGCTCAGGAAATTTGATGTT 840
|||
Db 22515 CGTAGTAATGGCGTCAATTTTATTTGCTTCAAGTTGTAGCGCTCAAGAGTTTGTATGTG 22574
|||
QY 841 ATCGGTACGTTAAACGCGCTGGAGCAACGCTTGAAGATATCATCACTCCGATGGAGCTT 900
|||
Db 22575 ATCAAAGTATGAATGAGCTTGAAGCTCGCTTAGAGGATATTCTGTGAGCCACTAGATCTG 22634
|||
QY 901 TCGATTATCGGCTGCGTGTGATGATGCCCGAGGTGAGCGCTGTTTCTACACTCGGCGTC 960
|||
Db 22635 TCAGTTATTGGCTGTAAAGTCAATGGCCAGGTGAAGCCAAAGACGACGATCGGTGTC 22694
|||
QY 961 ACCGCGGCAACAAGAAAAAGCGCTCTATGAAGATGGCGTGGCGCAAGACCGCTCTGGAC 1020
|||
Db 22695 GTGGGAGCCAGCCCAATAGTTTGTGTTTATAAATGTTGCAAAAAGCCATCTGATTGAT 22754
|||
QY 1021 ACAAACGATATGATCGACCGCTGGAGCAGCGATTCGTCGCAAAAGCCAGTCACTGAGTGGAC 1080
|||
Db 22755 ACCAAAAAATTTGATGATGAGATTGAGCGGATGTTGGCGTCTCAAGTCAAAAGACATTCAA 22814
|||
QY 1081 GAAGCGCTGCAATTCAGCTT 1101
Db 22815 GATAACGGGCAATGAGATT 22835
|||
RESULT 9
AAS90066
ID AAS90066 standard; cDNA; 1857 BP.
XX AAS90066;
AC
XX AAS90066;
DT 13-FEB-2002 (first entry)

Db 2555 TCGGCGACACGCTGCGCGTATCGCTGGCGCGCATCGGTCGAAGAGATCAAAAGTCGGTT 2614
QY TCGATATTTTGAATCGCTCGGTATCGTTTCGGAGGATCAACTTCATCGCTGCCCGA 814
Db TCGATATTTTGAATCGCTCGGTATCGTTTCGGAGGATCAACTTCATCGCTGCCCGA 2674
QY CTGTTGCGCTCAGGAATTTGATGTTATCGGTACGGTTAAACGGCTGGAGCAACGCTGG 874
Db CTGTTGCGCTCAGGAATTTGATGTTATCGGTACGGTTAAACGGCTGGAGCAACGCTGG 2734
QY AAGATATCATCATCGATCGATGCTTTCGATTCATCGCTCGGTGGAATGGCCAGGTG 934
Db AAGATATCATCATCGATCGATGCTTTCGATTCATCGCTCGGTGGAATGGCCAGGTG 2794
QY AGCGCTGGTTTCTACACTCGGCTCACCGCGCGCAACGAAGAAAGCGCTCTATGAAG 994
Db AGCGCTGGTTTCTACACTCGGCTCACCGCGCGCAACGAAGAAAGCGCTCTATGAAG 2854
QY ATGGCTGCGCAAGACCGTCTGGCAACAACGATATGATCGACCGCTGGAGCAGCA 1054
Db ATGGCTGCGCAAGACCGTCTGGCAACAACGATATGATCGACCGCTGGAGCAGCA 2914
QY TTGCTGCGAAAGCCAGTCAGCTG 1077
Db TTGCTGCGAAAGCCAGTCAGCTG 2937

RESULT 7

ABA92787
ID ABA92787 standard; DNA; 640681 BP.
AC ABA92787;
XX ABA92787;
DT 27-MAR-2002 (first entry)
XX Buchnera sp. genomic DNA SEQ ID NO:1.
DE Buchnera; cockroach-symbiotic bacterium; cockroach extermination;
KW circular; ds.
XX Buchnera sp.
OS Buchnera sp.
PN JP2001292771-A.
XX 23-OCT-2001.
XX 07-APR-2000; 2000JP-0107160.
PF 07-APR-2000; 2000JP-0107160.
PR 07-APR-2000; 2000JP-0107160.
XX (RIKA) RIKAGAKU KENKYUSHO.
PA WPI; 2002-126043/17.
XX A genomic DNA of cockroach-symbiotic bacterium -
PT Claim 1; Page 16-230; 237pp; Japanese.
PS

XX The present invention describes a gene (I) derived from Buchnera sp.
XX containing the DNA (a) or (b), (a) has a fully defined base pair
XX sequence selected from a table of sequences found in the Buchnera sp.
CC genomic DNA of ABA92787 given in the specification or is a DNA selected
CC from complementary DNA sequences, and (b) is a DNA which hybridises with
CC the DNA (a) and encodes a protein. Also described are: (1) a recombinant
CC vector (II) containing (i); (2) a transformant (III) containing (II);
CC (3) a genomic DNA of Buchnera sp. containing the sequence given in
CC ABA92787; (4) a plasmid derived from Buchnera sp. containing DNA (c) or
CC (d). (c) is a DNA containing a fully defined sequence given in ABA92788
CC or ABA92789 and (d) is a plasmid which hybridises with a DNA; and (5) a
CC method for the preparation of a protein in which (III) is cultured and
CC the expression protein of the objective protein is collected from the
CC resultant culture. The DNA is useful for developing agricultural
CC chemicals for exterminating cockroaches. The present sequence represents

CC the specifically claimed Buchnera sp. genomic DNA sequence, from the
CC present invention.
XX Sequence 640681 BP; 237522 A; 83822 C; 84757 G; 234580 T; 0 other;
SQ Query Match 36.1%; Score 403.6; DB 24; Length 640681;
Best Local Similarity 62.2%; Pred. No. 7.5e-106;
Matches 653; Conservative 0; Mismatches 394; Indels 3; Gaps 1;
QY 19 ATCAACGTAGAAAATCAACACGTAATTTACGTTTGGGAATGTCCGATTTGGCGATGGTGT 78
Db 314293 ATCAATAGAGAAAATCTGATCGTATTTATTTGGAAAAAGTCGATTTGGCAATAATGCG 314352
QY 79 CCATCGCGGTACAGTCCATGACCAATACGCTACGACAGAGCTGGAAGCAACGGTCAAT 138
Db 314353 CCAATATCAGTTCATCTATGACAAATCTCGTACTAATATCTCTGAAACATTAAT 314412
QY 139 CAAATCAAGCGCTCGAAGCGCTTGGCGCTGATATCGTCCGTGTATCCGTACCGACGATG 198
Db 314413 CAAATCTTAGAGTTACAAAAGTAGGAGTAGATATTTGTCGTATTTCTATACCAATTTA 314472
QY 199 GACGCGCAGAAAGCGTTTCAAATCTCAAAACAGCAGGTTAAACGTCGCGCTGGTGGCTGAC 258
Db 314473 AAAGCTGCAGAAATCATTCAAAGAAAATAAAAAACAAACAAATGTTCCATTTGATTGCGAT 314532
QY 259 ATCCACTTCGACTATCGCATTCGCGTTGAAAGTAGCGGATACGGCTCGCTGCTGCGT 318
Db 314533 ATACATTTTGATTACAGATTAGCTTTTACAAGTATAAATAATGGTGCAATTTGTTTAAAGA 314592
QY 319 ATTAACCCCTGGCAATATCGGTAATGAAGAGCGTATTTGCGCATGGTGTGACTGTGCGCGC 378
Db 314593 ATTAATCTGGGAATATTGGAATAAAGAAAGATACAGAATCATTTCTTACGCAAAA 314652
QY 379 GATAAAACATTTCCGATCCGTAATTTGGCTTAACGCGGATCCGCTGGAAGAAAGATCTGCAA 438
Db 314653 GATGAAAATATTCCAAATTCGTAATTTGGTAAATGCTGGATCTTTAGAAAAAGATATATTA 314712
QY 439 GAAAAGTAGTGGCAACCGACCGCGGAGGTTGCTGGAATCTGCCATCGCTCATGTTGAT 498
Db 314713 AAAAAATATAAATAACCTACTCCAGATGCATTTAGTAGAATACAGTATAGGCGATATTGAA 314772
QY 499 CATCTCGATCGCTGAACTTCGATCAGTTTCAAGTCAGCGTCAGAAAGCGCTCTGACGCTTC 558
Db 314773 TACTTTGATGCTTTAAATTTAAATCAATTTAAAGTTAGTTTAAAGCGCTCATGTTATTT 314832
QY 559 CTCGCTGTGAGCTTATCGTTTGTGCGCAACAGATCGATCAGCCGTTGCAATCTGGGG 618
Db 314833 TTAGCTATTGAATCATATCGCATGTTAGGAAAAAAATTAACAACCTTTGCAATATTGGC 314892
QY 619 ATCCCGAAGCGGTGTCGCGCAGCGCGGAGGTAATAATCCGCCATTTGGTTAGGCTG 678
Db 314893 ATAACTGAATCCGGTGGTTTAAAGGAATGGAACAGTTAAATCATCTATAGTATTGCTTTA 314952
QY 679 CTGCTGTCTGAAGGCATCGCGGACACGCTGCGCGTATCGCTGGCGCGCATCCGGTCGAA 738
Db 314953 TTATTTTAGAAGGCAATTTGGAGATACATACGAGTTTCAATAGCGGCACATCCAACTGAA 315012
QY 739 GAGATCAAAGTCGTTTCGATATTTTGAATTCGCTGCGTATCCGTTCCGAGGATCAAC 798
Db 315013 GAAGTAAAGTAGGTTATGACATTTTAAAGTTTATCTTTAAGAGCAAGAGGATTAAAT 315072
QY 799 TTCATCGCTCCCGACCTGTCGCGTCAGGAATTTGATTTGATTCGTTACGTTAACGGC 858
Db 315073 TTTATTTGCTGCTACTTTGTTCTAGACAAGAAATTTGATTAATTAACAGTAAATCAA 315132
QY 859 CTGAGAACCGCTCGGAAGATATCATCTCCGATGGAGCTTTTCGATTTATCGCTGCGGTG 918
Db 315133 CTAGAAAAAATCTAGAGAGATATCTCGACTCCATAGATGATCAATTTATTTGGTTGCGTT 315192
QY 919 GTGAATGCCCAGGTGAGGCGCTGGTTTCTACATCGCGCTCACCGCGCGCAACAGAAA 978
Db 315193 GTTAATGAATAGGTGAATCTAAAAATAGCAACTTTAGGTCTGACGAGGAAGTCATAGAAA 315252

QY 10 CAGGCTCCAAATCAACGTAGAAAAATCAACACGATTTTACGTTGGGAATGTCGCGATTGCG 69
DB 390972 CAGCCNACTATTAAAGCGTCGTGATCGACAAAAATTTATGTGGAAATGTACCAATTTGT 391031
QY 70 GATGTGCTCCATCCGCTGACGTCCATGACCAATACCGTACGACGACGTCGAAGCA 129
DB 391032 GGGGATGCGCTATTGCGGTGCAATCAATGACAAATACTCGACCACTGATGTGAAGCG 391091
QY 130 ACGTCAATCAATCAAGCGCTGGAACGGTTGGCGTGATATCGCTGATCCGTA 189
DB 391092 ACAGTTGCTCAAAATTAATCATTTAGAACGTTGTTGGCGATATTTCTGCTGATCTGTT 391151
QY 190 CCGACGATGAGCGCGCAGAGCGTTCAAACTCATCAACAGCAGGTAAACGTCGCGCTG 249
DB 391152 CCAACNATGATGCTCGGAAGCATTTAAACAAATTAACAAAGTGAATGTTCCGCTC 391211
QY 250 GTGGCTGACATCCACTTCGACTATCGATTCGCTGAAAGTACGCGAATACGCGCTCGAT 309
DB 391212 GTAGCAGATATTCATTTGACTATCGTATCGCGTTAAAGTCGCAATATGGAGTGGAT 391271
QY 310 TGTCTCGGTATTAAACCTGCAATATCGGTAAATGAAGCGTATTCGCTGCTGTTGAC 369
DB 391272 TGTTCAGTATCAATCTCGCAACATTTGCTGCGAAGATCGCGTCCGCTGCTGTTGAT 391331
QY 370 TGTGCGCGGATAAAAACATTCGATCCGTTATTTGGCGTTAAACGCGGATCGCTGGAATA 429
DB 391332 TGTGCGCGGAGAAAAATATTCGATTCGATTTGTTGTTGTTAAATGACAGCTCTTTAGAAAA 391391
QY 430 GATCTGCAAGAAAGTATGCGCAACGCGCGCGCGGTTGCTGGAATCTGCAATCGCT 489
DB 391392 GATTTGCAAGAAAAATATGCGCAACCAACGCGCAAGCGCTTTGTAAGATCCGCTGCT 391451
QY 490 CATGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTACGCGTGAAGCGTCT 549
DB 391452 CATGTAGAAATTTAGATCGCTTAACTTCGATCAGTTTAAAGTGAGCGTAAAGCGCTCC 391511
QY 550 GAGCTCTCTCGCTGTTGAGTCTTATCGTTTCTGCAAAACAGATCGATCAGCGCTG 609
DB 391512 GATGTATTTCTAGCGTTGAACTTATCGTTTACTGCTTAAGCAATTAACAGCGCTTTA 391571
QY 610 CATCTGGGATCAACGAACGCGTGTGCGCGCAGCGGGCAGTAAATCCGCGATTTGTT 669
DB 391572 CATTTAGGCATTAACAAGCAGGTGCGCAGCGCTGGTGCAGTAAATCTGCAGTGGGT 391631
QY 670 TTAGTCTGCTGCTGCTGAGGATCGCGCAGCGCTGCGGTGATGCTGCGCGCGCGAT 729
DB 391632 TTAGGAATGTTATTAGCTGAGGCGATTTGGCGATACACTACGCGCTCTCTTTGGCGCGAGAT 391691
QY 730 CCGGTCGAAGAGATCAAGTTCGTTTTCGATATTTTGAATTCGCTGCTATCCGTTCCGGA 789
DB 391692 CCTGTAGAGGAATCAAGTTCGTTTGTATTTTGAATTCCTTACGATTCGTTCAAGA 391751
QY 790 GGGATCAACTTCATCGCTGCCGACCTGTTTCGCGTCAGGAATTTGATGTTATCGGTACG 849
DB 391752 GGAATTAACCTTATTCGTTGCCCAACCTGTTCTCGCAAGATTTGATGTAATCGTACA 391811
QY 850 GTTAACGCGCTGAGCAACCGCTGGAAGATATCATCTCCGATGAGCGTTTCGATTC 909
DB 391812 GTAAATGCGCTAGAACAAACCGCTTGAAGATATTATTACCAATGATGATCTATTATC 391871
QY 910 GCGTGGTGTGATGCGCGAGGTGAGCGCTGTTTCTACACTCGCGCTCACCGCGCG 969
DB 391872 GGTGTGTGATGATGTTGCTGCGAGGCACTCGTCTCCGATCTCGCGTAACGGCGGT 391931
QY 970 AACAGAAAGCGCGCTCTATGAGATGGCGTGCGC---AAAGACCGTCTGGCAACAAC 1026
DB 391932 AACAAAAAGCGTTATTATCTTGACGGAGAGCGCAAAAGAGCGCTTTTGATACGAA 391991
QY 1027 GATATGATCGACGCTGGAAGCAGCGATTTCTGTCGGAAGCC 1068
DB 391992 GATATAGTGAACCAATTAGAAAGCAAAAATTTCTGTCGGAAGTTC 392033

RESULT 6
AAS73006
ID AAS73006 standard; cDNA; 5484 BP.
XX
AC AAS73006;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #8810.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
XX
DR P-PSDB; ABG08819.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity -
XX
PS Claim 1; SEQ ID No 8810; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic coding sequences of the invention.
XX
CC Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 5484 BP; 1153 A; 1384 C; 1583 G; 1364 T; 0 other;
Query Match 36.8%; Score 411.8; DB 23; Length 5484;
Best Local Similarity 97.1%; Pred. No. 4e-109;
Matches 430; Conservative 0; Mismatches 12; Indels 1; Gaps 1;
QY 635 GTGCGCGCAGCGGGCAGTAAATCCGCAATGCTTGTAGTCTGCTGCTCTGAAGGCA 694
DB 2496 GTGGGGCGCAGCGGAGTAAATCCCA-TGGTTAAGGTCTGCTGCTGAAGGCA 2554
QY 695 TCGGCGACACGCTGCGCGTATCGCTCGCGCGCGATCCGTCGAAGAGATCAAGTCCGTT 754

CC the same species, and to obtain nucleic acid homologues. gpE is also
CC used as or primers. The recombinant vectors are used in plant
CC transformation or transfection. gpE also act as markers capable of
CC detecting polymorphisms such as single nucleotide polymorphisms (SNPs).
CC gpE is also used to determine the level or pattern of expression of
CC the protein. The present sequence is Escherichia coli gpE gene.
XX
SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Query Match 100.0%; Score 1119; DB 24; Length 1119;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCATAACACAGGCTCAATTCACGTAGAAAATCAACAGCTATTACGTTGGGAATGTG 60
DB |||
1 ATGCATAACACAGGCTCAATTCACGTAGAAAATCAACAGCTATTACGTTGGGAATGTG 60

QY 61 CCAGTTGGGATGGTCTCCATCGCGCTACAGTCCATACCAATACGCGTACGACAGAC 120
DB |||
61 CCAGTTGGGATGGTCTCCATCGCGCTACAGTCCATACCAATACGCGTACGACAGAC 120

QY 121 GTCGAGCAACGGTCAATCAATCAAGGGCGCTGGAACGGCTTGGCGCTGATATCGTCCGT 180
DB |||
121 GTCGAGCAACGGTCAATCAATCAAGGGCGCTGGAACGGCTTGGCGCTGATATCGTCCGT 180

QY 181 GTATCCGTACCGACGATGAGCGGCGAGAGCGCTTCAAACTCATCAACAGCAGGTTAAC 240
DB |||
181 GTATCCGTACCGACGATGAGCGGCGAGAGCGCTTCAAACTCATCAACAGCAGGTTAAC 240

QY 241 GTGCCGCTGGTGGCTGACATCCACTTTCGATTCGCAATTCGCGTGAAGTAGCGGAATAC 300
DB |||
241 GTGCCGCTGGTGGCTGACATCCACTTTCGATTCGCAATTCGCGTGAAGTAGCGGAATAC 300

QY 301 GCGTCGATGCTGGTATTAACCTTGGCAATATCGGTAAATCAAGAGGTTATTCGATG 360
DB |||
301 GCGTCGATGCTGGTATTAACCTTGGCAATATCGGTAAATCAAGAGGTTATTCGATG 360

QY 361 GTGGTTGACTGTGCGCGCATAAACATTTCCGATCCGTTTAAACGCGGATCG 420
DB |||
361 GTGGTTGACTGTGCGCGCATAAACATTTCCGATCCGTTTAAACGCGGATCG 420

QY 421 CTGGAAGAAAGATCTGCAAGAAAGTATGCGAACCGACCGCGCGTGTGCGAATCT 480
DB |||
421 CTGGAAGAAAGATCTGCAAGAAAGTATGCGAACCGACCGCGCGTGTGCGAATCT 480

QY 481 GCGATGCTCATGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540
DB |||
481 GCGATGCTCATGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540

QY 541 AAAGCGCTGACGCTCTTCTCGCTGTTGAGTCTTATCGTTTGGCAAAACAGATCGAT 600
DB |||
541 AAAGCGCTGACGCTCTTCTCGCTGTTGAGTCTTATCGTTTGGCAAAACAGATCGAT 600

QY 601 CAGCGCTTCATCTGGGGATCACCGAAGCGGCTGTGCGCGCAGCGGGGACGTAATCC 660
DB |||
601 CAGCGCTTCATCTGGGGATCACCGAAGCGGCTGTGCGCGCAGCGGGGACGTAATCC 660

QY 661 GCCATTGTTAGTCTGCTGCTGCTGAAGGATCGGCGACACGCTCGCGGTATCGCTG 720
DB |||
661 GCCATTGTTAGTCTGCTGCTGCTGAAGGATCGGCGACACGCTCGCGGTATCGCTG 720

QY 721 GCGGCGATCCGCTCGAAGAGATCAAGTCGGTTTCGATATTTGAAATCGCTGCTATC 780
DB |||
721 GCGGCGATCCGCTCGAAGAGATCAAGTCGGTTTCGATATTTGAAATCGCTGCTATC 780

QY 781 CGTTCCGAGGATCAACTTCATTCGCTGCCGACCTGTTCCGCTCAGGAATTCATGTT 840
DB |||
781 CGTTCCGAGGATCAACTTCATTCGCTGCCGACCTGTTCCGCTCAGGAATTCATGTT 840

QY 841 ATCGGTACGGTTAAACGCGCTGGAGCAACCGCTGGAGATATCATCTCCGATGACGTT 900
DB |||
841 ATCGGTACGGTTAAACGCGCTGGAGCAACCGCTGGAGATATCATCTCCGATGACGTT 900

QY 901 TCGATTATCGGCTGGCTGTGATGATGCCAGGTGAGCGCTGTTTCTACACTCGCGTC 960
DB |||
901 TCGATTATCGGCTGGCTGTGATGATGCCAGGTGAGCGCTGTTTCTACACTCGCGTC 960

QY 961 ACCGCGGCAACAAGAAAGCGCTCTATGAGATGGCGTGGCGCAAGACCGTCTGGAC 1020
DB |||
961 ACCGCGGCAACAAGAAAGCGCTCTATGAGATGGCGTGGCGCAAGACCGTCTGGAC 1020

QY 1021 AACACGATATGATCGACCGCTGGAAGCACCGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080
DB |||
1021 AACACGATATGATCGACCGCTGGAAGCACCGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080

QY 1081 GAAGCGCTCGAATTCAGCTTCAGAGGTTGAAAATAA 1119
DB |||
1081 GAAGCGCTCGAATTCAGCTTCAGAGGTTGAAAATAA 1119

RESULT 5
AAT42063
ID AAT42063 standard; DNA; 1830121 BP.
XX
AC AAT42063;
XX
DT 14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
KW Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; ds.
XX
OS Haemophilus influenzae.
XX
PN WO9633276-A1.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US05320.
PR 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UJO) UNIV JOHNS HOPKINS.
PI
PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX WPI; 1996-485782/48.
DR
XX Haemophilus influenzae Rd genome recorded on computer readable
PT medium - useful for identifying commercially important nucleic acid
PT fragments by homology searching
XX
PS Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC This sequence represents the complete genome sequence of the bacterium
CC Haemophilus influenzae strain Rd. The invention relates to a computer
CC readable medium (CRM) having recorded upon it the complete H. influenzae
CC nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC sequence at least 99% identical to (I). By providing the full-length
CC genomic sequence in a computer readable form, it is possible to identify
CC commercially important nucleic acid fragments and expression modulating
CC fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC regulate the expression of a nucleic acid molecule. Vectors and altered
CC organisms comprising the predicted ORFs can be used to produce any of the
CC polypeptide fragments of the H. influenzae Rd genome.
SQ Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;

Query Match 55.6%; Score 622; DB 17; Length 1830121;
Best Local Similarity 74.8%; Pred. No. 3.2e-169;
Matches 794; Conservative 0; Mismatches 265; Indels 3; Gaps 1;

CC and (iii) to identify compounds that inhibit activity of gcpE, i.e.
CC potential antibacterial, antimycotic, antiparasitic or antiviral agents
CC for use in humans or animals, or antiviral, antiparasitic, fungicidal
CC or herbicidal agents for agriculture.
XX
SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;
Query Match 100.0%; Score 1119; DB 22; Length 1119;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCATTAACAGGCTCAATTCACCTAGAAATCAACACGATATTACGTTGGGAATGTG 60
Db 1 ATGCATAACAGGCTCAATTCACCTAGAAATCAACACGATATTACGTTGGGAATGTG 60
QY 61 CCGATTGGGATGGTCTCCATCGCGGTACAGTCCATGACCAATACGGGTACGACAGAC 120
Db 61 CCGATTGGGATGGTCTCCATCGCGGTACAGTCCATGACCAATACGGGTACGACAGAC 120
QY 121 GTCGAAGCAACGGTCAATCAAAATCAAGGCGCTGGAACGGTTGGCGCTGATATCGTCCGT 180
Db 121 GTCGAAGCAACGGTCAATCAAAATCAAGGCGCTGGAACGGTTGGCGCTGATATCGTCCGT 180
QY 181 GTATCGGTACCGACGATGGAACGGCGAGAGCGTTTCAAACTCATCAACAGCAGGTTAAAC 240
Db 181 GTATCGGTACCGACGATGGAACGGCGAGAGCGTTTCAAACTCATCAACAGCAGGTTAAAC 240
QY 241 GTGCCCGTGGTGGTGCATCTCACTTCGATATCGGATTCGATTCGATTCGATTCGATTCG 300
Db 241 GTGCCCGTGGTGGTGCATCTCACTTCGATATCGGATTCGATTCGATTCGATTCGATTCG 300
QY 301 GGGCTCGATTGCTCGGTATTAACCTGGCAATATCGGTAAATCAAGAGCGTATTCGCATG 360
Db 301 GGGCTCGATTGCTCGGTATTAACCTGGCAATATCGGTAAATCAAGAGCGTATTCGCATG 360
QY 361 GTGGTTGACTGTGCGCGGATATAAAACATTCGATTCGATTCGATTCGATTCGATTCG 420
Db 361 GTGGTTGACTGTGCGCGGATATAAAACATTCGATTCGATTCGATTCGATTCGATTCG 420
QY 421 CTGGAAGAAAGATCTGAAGAAAGATATGGGAACCGAGCGCGGCGGTTGCTGGAATCT 480
Db 421 CTGGAAGAAAGATCTGAAGAAAGATATGGGAACCGAGCGCGGCGGTTGCTGGAATCT 480
QY 481 GCCATCGCTATGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 540
Db 481 GCCATCGCTATGCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 540
QY 541 AAAGCGTCTGACGCTCTTCTCGCTGTTAGTCTTATGCTTATGCTTATGCTTATGCTTAT 600
Db 541 AAAGCGTCTGACGCTCTTCTCGCTGTTAGTCTTATGCTTATGCTTATGCTTATGCTTAT 600
QY 601 CAGCGGTGATCTGGGATACCGAAGCGCGGTGGTGGCGGACGCGGCGGAGTAAATCC 660
Db 601 CAGCGGTGATCTGGGATACCGAAGCGCGGTGGTGGCGGACGCGGCGGAGTAAATCC 660
QY 661 GCCATTGTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
Db 661 GCCATTGTTAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
QY 721 GCGGCGGATCCGCTGGAAGAGATCAAAAGTCCGTTTGCATATTTGAAATCCGCTGATC 780
Db 721 GCGGCGGATCCGCTGGAAGAGATCAAAAGTCCGTTTGCATATTTGAAATCCGCTGATC 780
QY 781 CGTTGCGAGGATCAATCTTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
Db 781 CGTTGCGAGGATCAATCTTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
QY 841 ATCGGTACCGTTAACCGCTGGAGCAACGCTCGAGATATCATCTCCGATGGACGTT 900
Db 841 ATCGGTACCGTTAACCGCTGGAGCAACGCTCGAGATATCATCTCCGATGGACGTT 900
QY 901 TCGATTATCGGCTCGCTGGTGAATGGCCACAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960
Db 901 TCGATTATCGGCTCGCTGGTGAATGGCCACAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960

Db 901 TCGATTATCGGCTCGCTGGTGAATGGCCACAGGTGAGGCGCTGGTTTCTACACTCGGCGTC 960
QY 961 ACCGCGGCAACAAGAAAGCGCTCTATGAAGATGGCTGCGCAAGACCGTCTGGAC 1020
Db 961 ACCGCGGCAACAAGAAAGCGCTCTATGAAGATGGCTGCGCAAGACCGTCTGGAC 1020
QY 1021 AACACAGATATGATCGACAGCTGGAAGCAGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080
Db 1021 AACACAGATATGATCGACAGCTGGAAGCAGCATTCGTGCGAAAGCCAGTCAGCTGGAC 1080
QY 1081 GAAGCGCTGGAATTCAGCTTCAGCAGGTTGAAAAATAA 1119
Db 1081 GAAGCGCTGGAATTCAGCTTCAGCAGGTTGAAAAATAA 1119
RESULT 4
AAD31203
ID AAD31203 standard; DNA; 1119 BP.
XX
AC AAD31203;
XX DT 31-MAY-2002 (first entry)
XX Escherichia coli gcpE gene.
DE DE
XX gcpE gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate;
KW transgenic plant; isoprenoid compound; tocopherol; isopentenyl diphosphate;
KW food; feed source; transfection; single nucleotide polymorphism; SNP;
KW oxidative stress tolerance; UV tolerance; transformation; plant; ds.
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..1119
FT /tag= a
FT /product= "Escherichia coli GCPE protein"
PN WO200212478-A2.
XX
PD 14-FEB-2002.
XX
PF 06-AUG-2001; 2001WO-US24335.
XX
PR 07-AUG-2000; 2000US-223483P.
XX
PA (MONS) MONSANTO TECHNOLOGY LLC.
XX
PI Boronat A, Campos N, Rodriguez-concepcion M, Rohmer M, Seeman M;
PI Valentin HE, Venkatesh TV, Venkatramesh M;
XX
XX WPI: 2002-227151/28.
DR P-PSDB; AAE19653.
XX
PT gcpE nucleic acid which is an essential gene of the methyl-D-erythritol
PT phosphate pathway, encoding a fully defined GCPE protein which is
PT useful for increasing levels of tocopherol substrates in plants -
XX
XX Claim 3; Page 117-119; 155pp; English.
XX
CC The invention relates to gcpE nucleic acid molecule, an essential gene
CC of methyl-D-erythritol phosphate (MEP) pathway that encodes rice,
CC Arabidopsis thaliana or Escherichia coli GCPE protein. gcpE is useful
CC for producing a transgenic plant such as Brassica campestris, B. napus,
CC canola, castor bean, coconut, cotton, crame, linseed, maize, mustard,
CC oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower,
CC or wheat with an increased isoprenoid (tocopherol) compound level. The
CC expression of GCPE protein in organisms increases the level of
CC tocopherol substrate such as isopentenyl diphosphate and dimethylallyl
CC diphosphate biosynthesis. Transgenic organisms overexpressing GCPE
CC protein can nutritionally enhance food and feed sources. Overexpression
CC of GCPE protein in transgenic plant may provide tolerance to stresses
CC e.g., oxidative stress tolerance such as to oxygen or ozone, UV
CC tolerance, etc. gcpE may be used to obtain nucleic acid molecules from

CC potential in higher organisms (mostly no orthologues were identified
CC in *Saccharomyces cerevisiae*). An antagonist or inhibitor of the
CC expression of an essential gene or of its function provides the key
CC for antibacterial therapy. The invention provides methods for
CC identifying such antagonists or inhibitors. These involve
CC contacting a bacterial cell comprising an essential gene with a
CC candidate antagonist or inhibitor, and testing whether contact leads
CC to cell growth inhibition and/or cell death. The method allows the
CC development of new broad spectrum antibiotics. A conditional mutant
CC of an essential gene can be used to induce a lethal phenotype in
CC bacteria for the analysis of surrogate markers.

XX Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;

Query Match		100.0%;	Score 1119;	DB 21;	Length 1119;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 1119;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	ATGCATAACCCAGGCTCCAAATTCACAGTAGAATAATCAACACGATATTTACGTTGGGAATGTG	60		
Db	1	ATGCATAACCCAGGCTCCAAATTCACAGTAGAATAATCAACACGATATTTACGTTGGGAATGTG	60		
QY	61	CCGATTGGCGATGGTCTCCATCGCGGTACAGTCATGACCAATACGGGTACGACAGAC	120		
Db	61	CCGATTGGCGATGGTCTCCATCGCGGTACAGTCATGACCAATACGGGTACGACAGAC	120		
QY	121	GTGGAAGCAACCGTCAATCAATCAAGCGCTGGAACGGTGGCGCTCATATCGTCGT	180		
Db	121	GTGGAAGCAACCGTCAATCAATCAAGCGCTGGAACGGTGGCGCTCATATCGTCGT	180		
QY	181	GTATCCGTACCGACGATGGACGGCGGAGAGCGTTCAAACTCATCAACACGAGGTTTAAAC	240		
Db	181	GTATCCGTACCGACGATGGACGGCGGAGAGCGTTCAAACTCATCAACACGAGGTTTAAAC	240		
QY	241	GTGCGCGTGGTGTGACATCACTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG	300		
Db	241	GTGCGCGTGGTGTGACATCACTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG	300		
QY	301	GGCGTCGATTGTCTGGGTATTAACCTCGCATATCGGTAAATGAAGAGCGTATTCGCATG	360		
Db	301	GGCGTCGATTGTCTGGGTATTAACCTCGCATATCGGTAAATGAAGAGCGTATTCGCATG	360		
QY	361	GTGGTTGACTGTGCGCGGATATAAATTCGATTCGATTCGATTCGATTCGATTCGATTCG	420		
Db	361	GTGGTTGACTGTGCGCGGATATAAATTCGATTCGATTCGATTCGATTCGATTCGATTCG	420		
QY	421	CTGGAAAAAGATCTGCAAGAAAGTATGCGGAACCGACCGCGAGCGTTGCGAATCT	480		
Db	421	CTGGAAAAAGATCTGCAAGAAAGTATGCGGAACCGACCGCGAGCGTTGCGAATCT	480		
QY	481	GCCATGCGTCATGTTGATCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG	540		
Db	481	GCCATGCGTCATGTTGATCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG	540		
QY	541	AAAGCGTCTGAGCTCTTCTCGCTGTGTAGTCTTATCGTTTCTGCGCAAAACAGATCCGAT	600		
Db	541	AAAGCGTCTGAGCTCTTCTCGCTGTGTAGTCTTATCGTTTCTGCGCAAAACAGATCCGAT	600		
QY	601	CAGCGCTTCATCTGGGGATCAACGAGCCGTTGGTGGCGCGAGCGGGCAGTAAATCC	660		
Db	601	CAGCGCTTCATCTGGGGATCAACGAGCCGTTGGTGGCGCGAGCGGGCAGTAAATCC	660		
QY	661	GCCATGTTTGTAGTCTGCTGTCTGAAGCATCGGACACGCTGGCGGTATTCGCTG	720		
Db	661	GCCATGTTTGTAGTCTGCTGTCTGAAGCATCGGACACGCTGGCGGTATTCGCTG	720		
QY	721	GCGGCGGATCCCGTCAAGAGATCAAAATCGGTTTTCGATATTTTGAATTCGCTGATC	780		
Db	721	GCGGCGGATCCCGTCAAGAGATCAAAATCGGTTTTCGATATTTTGAATTCGCTGATC	780		
QY	781	CGTTCGCGAGGATCAACTTCATCGCCTCCCGACCTGTCGCGTCAAGAAATTTGATGTT	840		
Db	781	CGTTCGCGAGGATCAACTTCATCGCCTCCCGACCTGTCGCGTCAAGAAATTTGATGTT	840		

QY	841	ATCGGTACGGTTAACGCGCTGGAGCAACGCCCTGGAGATATCATCTCCGATGGACGTT	900
Db	841	ATCGGTACGGTTAACGCGCTGGAGCAACGCCCTGGAGATATCATCTCCGATGGACGTT	900
QY	901	TCGATTATCGGCTCGGTGGTGAATGGCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC	960
Db	901	TCGATTATCGGCTCGGTGGTGAATGGCCAGGTGAGGCGCTGGTTTCTACACTCGGCGTC	960
QY	961	ACCGCGCGCAACAAGAAAAAGCGGCTCTATGAAGATGGCGTGGCAAGACCGTCTGGAC	1020
Db	961	ACCGCGCGCAACAAGAAAAAGCGGCTCTATGAAGATGGCGTGGCAAGACCGTCTGGAC	1020
QY	1021	AACAAACGATATGATCGACCGCTGGAAGACGCGCATTCGTGCGAAAGCCAGTCAGCTGGAC	1080
Db	1021	AACAAACGATATGATCGACCGCTGGAAGACGCGCATTCGTGCGAAAGCCAGTCAGCTGGAC	1080
QY	1081	GAAGCGGCTCGAATTTGACGTTTCAGCAGGTTGAAAAATAA	1119
Db	1081	GAAGCGGCTCGAATTTGACGTTTCAGCAGGTTGAAAAATAA	1119
RESULT 3			
ID	AAC82653		
XX	AAC82653 standard; DNA; 1119 BP.		
AC	AAC82653;		
DT	15-MAR-2001 (first entry)		
DE	E. coli gcpE DNA.		
KW	Isoprenoid; gcpE; yfgB; antimicrobial; transgenic plant; agriculture;		
KW	antimycotic; antiparasitic; antiviral; fungicidal; herbicidal; ds.		
OS	Escherichia coli.		
XX			
PN	WO200072022-A1.		
XX			
PD	30-NOV-2000.		
XX			
PF	20-MAY-2000; 2000WO-EP04592.		
XX			
PR	21-MAY-1999; 99DE-1023567.		
PR	21-MAY-1999; 99DE-1023568.		
XX			
PA	(JOMA/) JOMAA H.		
XX			
PI	Jomaa H;		
XX			
DR	WPI; 2001-025196/03.		
DR	P-PSDB; AAB45692.		
XX			
PT	Incorporating gcpE and yfgB genes into viruses and cells, for		
PT	increasing isoprenoid content and identifying e.g. antimicrobial		
PT	agents, comprises using DNA sequences from bacteria or parasites		
XX			
PS	Claim 3; Page 13-15; 36pp; German.		
XX			
CC	This invention describes a novel method for incorporating gcpE and yfgB		
CC	genes into viruses and cells for increasing isoprenoid content and		
CC	identifying e.g. antimicrobial agents, comprises using DNA sequences (I)		
CC	from the gcpE or yfgB genes of bacteria or parasites or DNA sequences		
CC	(II) which hybridize to the specified genes or encode a plastid protein		
CC	with the same biological activity as those encoded by the genes. The		
CC	invention also describes (i) plant cells containing (i) or (ii); (2)		
CC	transformed plant cells, and transgenic plants regenerated from them,		
CC	that contain (I) or (II); (3) determining the enzymatic activity of a		
CC	gcpE protein; or (4) screening compounds (A) that have antimycotic,		
CC	antiparasitic or antiviral activity in humans or animals or antiviral,		
CC	antiparasitic, fungicidal or herbicidal activity in plants. (i) and (ii)		
CC	are used: (i) to increase the isoprenoid levels in viruses and cells;		
CC	(ii) for determining the enzymatic activity of gcpE and yfgB proteins;		

DR WPI; 2000-687048/67.
XX Identifying antibacterial compounds, comprises identifying an
PT antagonist or inhibitor of the expression of a gene encoding a
PT polypeptide essential for bacterial growth or survival -
XX
PS Claim 1; Fig 1; 75pp; English.
XX
CC The present invention relates to antagonists and inhibitors of 24
CC bacterial genes and proteins. The proteins are thought to be essential
CC for growth in several species of bacteria (including *S. pneumoniae*, *B.*
CC *burgdorferi*, *H. influenzae* and *H. pylori*). The proteins and coding
CC sequences shown in the specification can be used to identify antagonists
CC and inhibitors which can be used in disease treatment and pesticides. In
CC particular, they can be used against *M. tuberculosis*. The present
CC sequence is one of the genes of the invention.
XX
SQ Sequence 1119 BP; 272 A; 279 C; 317 G; 251 T; 0 other;
Query Match 100.0%; Score 1119; DB 21; Length 1119;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCATAACACAGGCTCCAAATTCACGCTAGAAAATCAACAGCTATTACGTTGGGAATGTG 60
DB 1 ATGCATAACACAGGCTCCAAATTCACGCTAGAAAATCAACAGCTATTACGTTGGGAATGTG 60
QY 61 CCGATTGGCGATGGTGTCCCATCGCGGTACGTCATGACCAATACGCGTACGACAGAC 120
DB 61 CCGATTGGCGATGGTGTCCCATCGCGGTACGTCATGACCAATACGCGTACGACAGAC 120
QY 121 GTGGAAGCAACAGGTCATCAATCAATCAAGCGCTGGAACGCGTGTGCGCTGATATCGTCCGT 180
DB 121 GTGGAAGCAACAGGTCATCAATCAATCAAGCGCTGGAACGCGTGTGCGCTGATATCGTCCGT 180
QY 181 GTATCCGTACCGAGTGGACGCGGAGAAAGCGTTCAAACTCATCAACAGCAGGTTAAC 240
DB 181 GTATCCGTACCGAGTGGACGCGGAGAAAGCGTTCAAACTCATCAACAGCAGGTTAAC 240
QY 241 GTGCGCGTGTGGTGCATCCACTTCGATTCGCAATCGGATTCGGCTGAAAGTAGCGGAATAC 300
DB 241 GTGCGCGTGTGGTGCATCCACTTCGATTCGCAATCGGATTCGGCTGAAAGTAGCGGAATAC 300
QY 301 GGGTCGATTTGTCGCGTATTAACTCCCTGGCAATATCGTAAATGAAGAGCGTATTTCGCATG 360
DB 301 GGGTCGATTTGTCGCGTATTAACTCCCTGGCAATATCGTAAATGAAGAGCGTATTTCGCATG 360
QY 361 GTGTTGACCTGTGCGCGGATATAAAACATTCGATCCGTTATGCGGTTAAACGCGGATCG 420
DB 361 GTGTTGACCTGTGCGCGGATATAAAACATTCGATCCGTTATGCGGTTAAACGCGGATCG 420
QY 421 CTGGAATAAGATCTGCAAGAAAGATATGGCAACCGAGCGCGGCGTTGCTGGATCT 480
DB 421 CTGGAATAAGATCTGCAAGAAAGATATGGCAACCGAGCGCGGCGTTGCTGGATCT 480
QY 481 GCATGCGTCACTGTTGATCATCTGATCGCTGAACCTTCGATCAGTTCAAGTCAGCGTG 540
DB 481 GCATGCGTCACTGTTGATCATCTGATCGCTGAACCTTCGATCAGTTCAAGTCAGCGTG 540
QY 541 AAAGCGCTCAGCGCTTCTCCTCGCTGTTGAGTCTTATCGTTGCTGGCAAAACAGATCGAT 600
DB 541 AAAGCGCTCAGCGCTTCTCCTCGCTGTTGAGTCTTATCGTTGCTGGCAAAACAGATCGAT 600
QY 601 CAGCGTTGATCTGGGGATCAACGAAGCGCGTGTGTCGCGCAGCGGGGAGTAAATCC 660
DB 601 CAGCGTTGATCTGGGGATCAACGAAGCGCGTGTGTCGCGCAGCGGGGAGTAAATCC 660
QY 661 GCCATTGGTTAGTCTGCTGCTGCTGAAGGATCGCGCACACGCTGCGCGTATCGCTG 720
DB 661 GCCATTGGTTAGTCTGCTGCTGCTGAAGGATCGCGCACACGCTGCGCGTATCGCTG 720
QY 721 GCGGCCGATCGGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTCGCTATC 780
DB 721 GCGGCCGATCGGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTCGCTATC 780
DB 721 GCGGCCGATCGGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTCGCTATC 780
QY 781 CGTTCCGAGGATCAACTTCATCGCTGCCGACCTGTTCCGCTCAGGAAATTTGATGTT 840
DB 781 CGTTCCGAGGATCAACTTCATCGCTGCCGACCTGTTCCGCTCAGGAAATTTGATGTT 840
QY 841 ATCGGTACGTTTAAACGCGCTGGAGCAACGCTTGAAGATATCATCATCCGATGGACGTT 900
DB 841 ATCGGTACGTTTAAACGCGCTGGAGCAACGCTTGAAGATATCATCATCCGATGGACGTT 900
QY 901 TCGATTATCGGCTGGTGTGTAATGCCAGGTGAGGCGCTGTTTCTACACTCGCGCTC 960
DB 901 TCGATTATCGGCTGGTGTGTAATGCCAGGTGAGGCGCTGTTTCTACACTCGCGCTC 960
QY 961 ACCGCGGCAACAAGAAAGCGCTCTATGAAGATGGCGTGCAGAAACCGTCTGGAC 1020
DB 961 ACCGCGGCAACAAGAAAGCGCTCTATGAAGATGGCGTGCAGAAACCGTCTGGAC 1020
QY 1021 AACACGATATGATCGACCGCTGGAAGCACGCTTCGTGCGAAAGCCAGTCAGCTGGAC 1080
DB 1021 AACACGATATGATCGACCGCTGGAAGCACGCTTCGTGCGAAAGCCAGTCAGCTGGAC 1080
QY 1081 GAAGCGCTCGAATTGACGCTTCCAGAGGTTGAAAAATAA 1119
DB 1081 GAAGCGCTCGAATTGACGCTTCCAGAGGTTGAAAAATAA 1119
RESULT 2
AAA88705
ID AAA88705 standard; DNA; 1119 BP.
XX
AC AAA88705;
XX
DT 05-FEB-2001 (first entry)
XX
DE E. coli FUN essential gene gcpe.
XX
KW FUN gene; gcpe gene; essential gene; antibacterial; antibiotic;
KW screening; infection; therapy; antagonist; surrogate marker; ds.
XX
OS Escherichia coli.
XX
PN EP1043403-A1.
XX
PD 11-OCT-2000.
XX
PF 09-APR-1999; 99EP-0107031.
XX
PR 09-APR-1999; 99EP-0107031.
XX
PA (GPCG-) GPC GENOME PHARM CORP AG.
XX
PI Not given;
XX
WPI; 2000-640125/62.
XX
PT Identifying antagonists of the expression of gene encoding bacterial
PT growth polypeptide useful for treating bacterial infections or
PT diseases, by evaluating transcription of the gene in the presence of
PT test molecule -
XX
PS Claim 1; Page 25; 55pp; English.
XX
CC The present sequence is that of the Escherichia coli FUN gene
CC gcpe, which encodes a protein that is essential for bacterial
CC growth or survival. gcpe is 1 of 22 E. coli genes (see
CC AAA88692-713) identified as being essential (there is no deletion
CC genotype). These 22 genes fulfill criteria for being attractive
CC antibacterial targets: hypothetical open reading frames coding for
CC essential functions (mutation is lethal for growth in rich media);
CC broad conservation (orthologues are present in a wide range of
CC bacteria, including Haemophilus influenzae, Streptococcus pneumoniae,
CC Helicobacter pylori and Borrelia burgdorferi); and low toxicity

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 11:58:06 ; Search time 358 Seconds
(without alignments)
8437.629 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119

Sequence: 1 atgcataacaggctccaat.....ttcagcaggtgaaaaataa 1119

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_19Jun03.*

- 1: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1119	100.0	1119	21 AAA95478	E. coli essential
2	1119	100.0	1119	21 AAA88705	E. coli FUN essent
3	1119	100.0	1119	22 AAC82653	E. coli gcpe DNA
4	1119	100.0	1119	24 AAD31203	Escherichia coli g
5	622	55.6	1830121	17 AAT42063	Haemophilus influe
6	411.8	36.8	5484	23 AAS73006	DNA encoding novel
7	403.6	36.1	640681	24 ABA92787	Buchnera sp. genom
8	382.6	34.2	33140	22 AAF28536	Genomic fragment #

9	356	31.8	1857	23 AAS90066	DNA encoding novel
10	306.2	27.4	349980	24 ABQ81842	Bifidobacterium lo
11	293.4	26.2	1083	24 ABK74401	Bacillus lichenifo
12	254.6	22.8	6157	24 ABQ70939	Listeria monocytog
13	251.4	22.5	1107	24 ABQ67966	Listeria monocytog
14	251.4	22.5	1107	24 ABQ69993	Listeria monocytog
15	251.4	22.5	2944528	24 ABA03041	Listeria monocytog
16	244	22.5	35829	23 AAS59573	Propionibacterium
17	243.8	21.8	1161	25 ABZ71126	Mycobacterium tube
18	243.8	21.8	1164	22 AAH52056	Mycobacterium tube
19	243.8	21.8	4403765	22 AAI99683	Mycobacterium tube
20	243.8	21.8	4411529	22 AAI99682	Mycobacterium tube
21	243.6	21.8	994	24 ABQ68324	Listeria monocytog
22	239.4	21.4	1134	22 AAH67170	C glutamicum codin
23	239.4	21.4	1179	25 ACA00649	C glutamicum deri
24	239.4	21.4	349980	22 AAH68530	C glutamicum codin
25	239.4	21.4	349980	22 AAH68531	C glutamicum codin
26	211.8	18.9	1180	19 AAX14017	H. pylori GHPO 76
27	136.4	12.2	507	21 AAC75580	Human ORF1443 cDNA
28	136.4	12.2	507	24 ABN76496	Helicobacter pylor
29	102	9.1	435	24 ABX66784	DNA encoding novel
30	88	7.9	1152	23 AAS88480	Polynucleotide seq
31	87.2	7.8	2535	20 AAX20674	Sequence encoding
32	72.4	6.5	4467	10 AAN92428	Zea mays partial g
33	70.6	6.3	584	24 AAD31220	Zea mays partial g
34	68.4	6.1	596	24 AAD31222	Zea mays partial g
35	68.4	6.1	670	24 AAD31221	Zea mays partial g
36	58.8	5.3	1263	25 ABZ39756	N. gonorrhoeae nuc
37	58.8	5.3	60873	21 AAA81469	Neisseria meningit
38	58.8	5.3	349980	21 AAF21610	N. meningitidis B
39	58.8	5.3	1437668	21 AAA81490	Complete genome se
40	58.8	5.1	348	24 ABK78811	Bacillus clausii g
41	53.4	4.8	1038602	20 AAZ01425	Zea mays partial g
42	52.6	4.7	293	24 AAD31224	Zea mays partial g
43	52.6	4.7	456	24 AAD31225	Zea mays partial g
44	52.6	4.7	504	16 AAQ93779	Cross-reactive all
45	50.8	4.5	1398	20 AAX20609	Polynucleotide seq

ALIGNMENTS

RESULT 1

AAA95478
ID AAA95478 standard; DNA; 1119 BP.

XX AAA95478;

XX 27-FEB-2001 (first entry)

XX E. coli essential gene gcpe.

DE Bacterial growth; inhibitor; ygbB; yfhC; yacE; ychB; yefJ; yrfI;
KW yggJ; yjeE; ylaO; yrdC; yhcC; ygbE; ybeY; gcpE; kdtB; pfs; ycaJ;
KW H1808; yeaA; yagF; b1983; yidD; yceG; yjbc; antibacterial compound;
KW H. influenza; S. pneumoniae; H. pylori; B. burgdorferi; pesticide;
KW M. tuberculosis; antibiotic; ds.
XX Escherichia coli.
OS
XX
XX WO200061793-A2.
XX
XX PD 19-OCT-2000.
XX
XX PF 07-APR-2000; 2000WO-EP03135.
XX
XX PR 09-APR-1999; 99EP-0107031.
XX
XX PR 04-FEB-2000; 2000EP-0102111.
XX
XX PA (GPCB-) GPC BIOTECH AG.
XX
XX PI Loferer H, Jacobi A;


```

; GENERAL INFORMATION:
; APPLICANT: Boronat, Albert;
; APPLICANT: Campos, Narciso;
; APPLICANT: Rodriguez-Concepcion, Manuel;
; APPLICANT: Rohmer, Michel;
; APPLICANT: Seeman, Myriam;
; APPLICANT: Valentin, Henry E.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; APPLICANT: Venkatesh, Mylavaram;
; TITLE OF INVENTION: Methy1-D-Erythritol Phosphate Pathway Genes
; FILE REFERENCE: 16516.107/35-21(51897)US
; CURRENT APPLICATION NUMBER: US/09/921,992
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/223,483
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 22
; LENGTH: 670
; TYPE: DNA
; ORGANISM: Zea mays
US-09-921-992-22

```

```

Query Match      6.1%; Score 68.4; DB 9; Length 670;
Best Local Similarity 56.1%; Pred. No. 2.4e-12;
Matches 129; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy 595 ATCGATCAGCCGTTGCATCTGGGGATCACCGAAGCCGGTGGTGGCGCGGCGGCGTA 654
Db 223 ATGATTATCCCTCCATCTAGGGGTACCGAAGCCGGGATGGGAATATGCGCGATT 282

Qy 655 AAATCCGCCATTGGTTTAGTCTGCTGTCTGAAGGCATCGCGCACACGCTGCGCGTA 714
Db 283 AAATCCACCCTGGCATTGCACCCCTTTAGCTGATGGCATTGGCGATCTATCCGGTA 342

Qy 715 TCGCTGGCGGCCGATCCGGTCGAAGAGATCAAGTGGTTTCGATATTTTGAATCGCTG 774
Db 343 TCCCTCACCGAAGCCCCCGAAAAAGAAATTCGCGTTTGTCTACAGCATTTCTCCAGGGCGTG 402

Qy 775 CGTATCGTTTCGCGAGGGATCAACTTCATCGCCTGCCGACCTGTTCCGG 824
Db 403 GGTTTGGGAAACCATTGGTGAATATGTGGCCTGTCCTTCTCTGGCCG 452

```

Search completed: November 23, 2003, 17:14:25
Job time : 433 secs

Best Local Similarity 56.9%; Pred. No. 9.8e-39; Matches 265; Conservative 0; Mismatches 201; Indels 0; Gaps 0;									
QY	43	ATTTACGTTGGCGAATGTCGCGATGGCGATGGCTCCCATCCGCTACAGTCACATGACC	102						
Db	233	ATCAAAATCGGGATGTTAAAGTTGGGAAAAGAAATCCCGTTGTGATTCATCGATGATT	292						
QY	103	AATACGCGTACGACAGACGTCGAAGCAACGCTCAATCAATCAAGCGCTGGAAACGCGTT	162						
Db	293	AACACGMAACTCGGATGTAGAACGAGTGTACGGCAATTTGGATTGGGAAGACC	352						
QY	163	GGCGTGATATCGTTCGCTGTATCCGTAACGACGATGGAACGCGACGAAGGTTCAAACTC	222						
Db	353	GGCTGTGAATTGGTTCGTATGACGATCAATACAAAAGAGCGCTATGGCAATTCGCCGC	412						
QY	223	ATCAACACGACGTTAACGTCGCGCTGGTGCCTGACATCCACTTCGACTATCGATTGG	282						
Db	413	ATTAAAGAGAGTTTATATTCCTTTGGTACGGATATTCATTTGATTATCGACTTGGC	472						
QY	283	CTGAAAGTAGCGGAATACGCGCTCGATTGTCGTATTAAACCTGGCAATATCGGTAAT	342						
Db	473	TTATTGGCGATAGAACGGAATTGACAAATTACGATTAATCCCGAAATATCGGTTCT	532						
QY	343	GAAGAGCGTATTCGATGCTGGTGTGACTGTGCGCGCATAAAAACATTCGATCCGTAAT	402						
Db	533	GAAGAAATATTCGTTGGTTGCGGAAGCGGCAAAAAGGAAGAGAAATTCGGATTCTGTA	592						
QY	403	GGCGTTAACCGCGATCGCTGGAAAGATCTCAAGAAAGATATGGCGAACCGCGCG	462						
Db	593	GGAGTCAATGACGTTCTTTTGGAAAAACATATCTTGGAAAAATATGGAGCGGNAACGG	652						
QY	463	CAGCGCTGTGCTGGAATCTGCATCTGCATGCTGTTGATCATCTCGATC	508						
Db	653	GATGCTNTGTCAAAAGCGCTAATGATCATGTAAAACCTGTTGGAAC	698						
RESULT 13									
US-09-921-992-21/c									
; Sequence 21, Application US/09921992									
; Patent No. US20020069426A1									
; GENERAL INFORMATION:									
; APPLICANT: Boronat, Albert;									
; APPLICANT: Campos, Narciso;									
; APPLICANT: Rodriguez-Concepcion, Manuel;									
; APPLICANT: Rohrer, Michel;									
; APPLICANT: Seeman, Myriam;									
; APPLICANT: Valentin, Henry E.;									
; APPLICANT: Venkatesh, Tyamagondlu V.;									
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes									
; FILE REFERENCE: 16516.107/35-21(51897)US									
; CURRENT APPLICATION NUMBER: US/09/921,992									
; CURRENT FILING DATE: 2001-08-06									
; PRIOR APPLICATION NUMBER: US 60/223,483									
; PRIOR FILING DATE: 2000-08-07									
; NUMBER OF SEQ ID NOS: 85									
; SEQ ID NO 21									
; LENGTH: 584									
; TYPE: DNA									
; ORGANISM: Zea mays									
US-09-921-992-21									
Query Match 6.3%; Score 70.6; DB 9; Length 584;									
Best Local Similarity 52.5%; Pred. No. 4e-13;									
Matches 179; Conservative 0; Mismatches 159; Indels 3; Gaps 1;									
QY	595	ATCGATCAGCGTTGCATCTGGGGATCACCGAAGCGGTGGTCGCGCAGCGGGCAGTA	654						
Db	459	ATGGATTATCCCTCCATCTAGGGGTACCGAAGCGGGATGGGAATATGCCCAT	400						
QY	655	AAATCCGCCATGTTTGGTCTGCTGTCTGTAAGGCGATCGCGACACGCTGCGCGTA	714						
Db	399	AAATCCACCGCTGGCATTGCCACCCCTTTAGCTGATGGCATTTGGCGATATATCGGGTA	340						
RESULT 15									
US-09-921-992-22									
; Sequence 22, Application US/09921992									
; Patent No. US20020069426A1									
Query Match 6.1%; Score 68.4; DB 9; Length 596;									
Best Local Similarity 56.1%; Pred. No. 2.3e-12;									
Matches 129; Conservative 0; Mismatches 101; Indels 0; Gaps 0;									
QY	595	ATCGATCAGCGTTGCATCTGGGGATCACCGAAGCGGTGGTCGCGCAGCGGGCAGTA	654						
Db	223	ATGGATTATCCCTCCATCTAGGGGTACCGAAGCGGGATGGGAATATGCCCAT	282						
QY	655	AAATCCGCCATGTTTGGTCTGCTGTCTGTAAGGCGATCGCGACACGCTGCGCGTA	714						
Db	283	AAATCCACCGCTGGCATTTGCCACCCCTTTAGCTGATGGCAATGTCGCTGCTGTTGCGGTA	342						
QY	715	TCGCTGGGGCGCGATCCGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTG	774						
Db	343	TCCTTCACGAAGCGCCCGAAAAAGAAATTCCTGTTTGTACAGCATTTCTCAGGGCTG	402						
QY	775	CGTATCCGTTCCGAGGGATCAACTTCATTCGCTGCCCGACCTGTTTCGGG	824						
Db	403	GGTTTGGGAAAAACCATGCTGGAATATGTCGCTGCTGCTCTCTGTTGGCGG	452						
RESULT 14									
US-09-921-992-23									
; Sequence 23, Application US/09921992									
; Patent No. US20020069426A1									
; GENERAL INFORMATION:									
; APPLICANT: Boronat, Albert;									
; APPLICANT: Campos, Narciso;									
; APPLICANT: Rodriguez-Concepcion, Manuel;									
; APPLICANT: Rohrer, Michel;									
; APPLICANT: Seeman, Myriam;									
; APPLICANT: Valentin, Henry E.;									
; APPLICANT: Venkatesh, Tyamagondlu V.;									
; APPLICANT: Venkatesh, Tyamagondlu V.;									
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes									
; FILE REFERENCE: 16516.107/35-21(51897)US									
; CURRENT APPLICATION NUMBER: US/09/921,992									
; CURRENT FILING DATE: 2001-08-06									
; PRIOR APPLICATION NUMBER: US 60/223,483									
; PRIOR FILING DATE: 2000-08-07									
; NUMBER OF SEQ ID NOS: 85									
; SEQ ID NO 23									
; LENGTH: 596									
; TYPE: DNA									
; ORGANISM: Zea mays									
; FEATURE:									
; NAME/KEY: unsure									
; LOCATION: (1..596)									
; OTHER INFORMATION: unsure at all n locations									
US-09-921-992-23									
Query Match 6.1%; Score 68.4; DB 9; Length 596;									
Best Local Similarity 56.1%; Pred. No. 2.3e-12;									
Matches 129; Conservative 0; Mismatches 101; Indels 0; Gaps 0;									
QY	595	ATCGATCAGCGTTGCATCTGGGGATCACCGAAGCGGTGGTCGCGCAGCGGGCAGTA	654						
Db	223	ATGGATTATCCCTCCATCTAGGGGTACCGAAGCGGGATGGGAATATGCCCAT	282						
QY	655	AAATCCGCCATGTTTGGTCTGCTGTCTGTAAGGCGATCGCGACACGCTGCGCGTA	714						
Db	283	AAATCCACCGCTGGCATTTGCCACCCCTTTAGCTGATGGCAATGTCGCTGCTGTTGCGGTA	342						
QY	715	TCGCTGGGGCGCGATCCGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTG	774						
Db	343	TCCTTCACGAAGCGCCCGAAAAAGAAATTCCTGTTTGTACAGCATTTCTCAGGGCTG	402						
QY	775	CGTATCCGTTCCGAGGGATCAACTTCATTCGCTGCCCGACCTGTTTCGGG	824						
Db	403	GGTTTGGGAAAAACCATGCTGGAATATGTCGCTGCTGCTCTCTGTTGGCGG	452						

PRIOR FILING DATE: 1999-03-26
PRIOR APPLICATION NUMBER: 60/134,093
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/134,092
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: 60/165,124
PRIOR FILING DATE: 1999-11-12
PRIOR APPLICATION NUMBER: 60/165,086
PRIOR FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 292
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 110
LENGTH: 1164
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
US-09-712-363-110

Query Match 21.8%; Score 243.8; DB 10; Length 1164;
Best Local Similarity 53.5%; Pred. No. 3.3e-72;
Matches 534; Conservative 0; Mismatches 462; Indels 3; Gaps 1;

QY 49 GTTGGGATGCGGATGGGATGGTCTCCATCGCCGTACAGTCCATGACCAATACG 108
DB 76 GTGGCAACGTGCGGCGTGGGAGTGACCATCCGTCTCGGTGCAATCGATGTGCCACC 135
QY 109 CGTACGACAGACGTGGAAGCAACGGTCAATCAATCAAGCGCTGGAACCGTTGGCGCT 168
DB 136 ABAACCCACAGCTCACTGACATGCAATGCAACAATGCCAGTACCAGCGCGGATGC 195
QY 169 GATATCGTCCGTATCCGTACCGACGATGCGGCGGAGAGCGTTCAAACTCATCAA 228
DB 196 GACATCGTGGGTGGCTGCCCGCGCAGGAGAGCGCCGCGTGGCGAGATCGCC 255
QY 229 CAGCAGTTAACGTGCGGTGGTGGTGCATCCATCCATTCGATTCGATTCGCTGAAA 288
DB 256 CGCACAGCAGATCCCGGTAGTGGGAGACATATTCAGCGCGCTACATATTGCC 315
QY 289 GTAGCGGAATACGCGGCGATTGCTGCGTATTAACCCCTGGCAATATC---GGAATGAA 345
DB 316 GCATCGACGCTGATGTCGCGGTGGGTGCGGCGGATCAACCGGCAACATCAAGAGTTGAC 375
QY 346 GAGCGTATTCGATGTTGTTGATGTCGCGCGGATCAAAACATTCGATTCGATTTGGC 405
DB 376 GCGCGGTGGTGAAGTTCGCAAGCGGCGGTGGCGCGGATCCGATCCGAATCGGT 435
QY 406 GTTAAACCGCGATCGCTGGAAGAGTCTCAAGAAAGTATGCGCAACGACGCGCGAG 465
DB 436 GTCAACCGCGGTTGCTGGAACAAACGGTTTCATGGAAGATGTCGCAAGCACGCGCGAG 495
QY 466 GCGTTGCTGAATCTGCATGCGTCTGTTGATCATCTCGATCGCTGAACTTCGATCAG 525
DB 496 GCGCTGTTGAGTGGCGCTGTTGGAGGCTTCGCTTTGAGAGATGCGCTTCGTTGAC 555
QY 526 TTCAAAAGTACGCTGGAAGCGTCTGACGTCTTCTCGCTGTGAGTCTTATCGTTTGGTG 585
DB 556 ATCAAGATCAGGCTCAAGCAACGACCGCGGTGATGTTGCGCGCTACGAGTGCIT 615
QY 586 GCAAAACAGATCATAGCGGTTGATCTGGGATCACGAAGCGGTGTTGCGCGAGC 645
DB 616 GCTGCAAGTGCAGTACCACTGCACTGCGTGTACCGAGGCGCGCGCTTCTTCCAG 675
QY 646 GGGGAGTAAATCCGCAATGTTGTTAGTCTGCTGCTGCTCAAGCATCGCGCACAG 705
DB 676 GGCAACATCAAGTCCGCGTTGCTTCGCGCGGTTGCTTCGCGGCGATAGGCGACAC 735
QY 706 CTGCGGTATCGCTGGCGGCGATCCGCTGGAAGATCAAAAGTTCGATTTTGGT 765
DB 736 ATCGGCTGCTGTTGCTGCGCGCGCTGCGAGAGTCAAGTGGGCAATCAGGTTCTC 795
QY 766 AATCGCTGCTATCGTTGCGAGGATCACTTATCATCCCTGCGCGACCTTTCGCGT 825
DB 796 GAGTCGTTGAACCTGGCGCGGCTTCGCTCGAGATCGTGTCTTGGCGCGCTGCTGCG 855

QY 826 CAGGAATTTGATGTTATCGGTAAACGCGCTGAGCAACGCTGGAAGATATCATC 885
DB 856 GCGCAAGTCGAGCTTACACCCCTGCCAACAGGTAACCGCGGCTGATGGTCTCGAT 915
QY 886 ACTCCGATGAGCGTTTCGATTTATCGGCTCGGTGATGAATGGCCAGGTGAGGCGTGGTT 945
DB 916 GTGCCCTTTCGGGTGGCGGCTGATGGGTTGTCGTCATGTTGGTCCGGTGAAGCACGTCGAG 975
QY 946 TCTACACTCGGCGTCAACCGCGGCAACAGAAAGCGGCTCTATGAAGATGCGGTGCGC 1005
DB 976 GCGACCTGGGCGTGGCGTCCGGCAACGCAAGGTTCAGATCTTTTACGGGCGAAGTG 1035
QY 1006 AAAGACCGTCTGGCAACAAACGATATGATCGACCGCTG 1044
DB 1036 ATCAAGACCGTGGCGAAGCACAGATCGTCGAGACGCTG 1074

RESULT 9
US-09-738-626-2205
Sequence 2205, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 2205
LENGTH: 1134
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-2205

Query Match 21.4%; Score 239.4; DB 10; Length 1134;
Best Local Similarity 54.1%; Pred. No. 1e-70;
Matches 533; Conservative 0; Mismatches 446; Indels 6; Gaps 2;

QY 22 CAACGTAGAAAAATCAACAGTATTTACGTTGGGAATGTCGCGATGGCGATGGTGTCC 81
DB 10 CCAGTGTGTAACACGCAACTCATGTGCGCAAGTGGCGTGGTTCGATCACCG 69
QY 82 ATCGCGGTACAGTCCATGACCAATACGCGTACGACAGAGTCAAGCAACGGTCAATCAA 141
DB 70 ATTTCCGTCAGTCAGTACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 129
QY 142 ATCAAGCGCTGGAACGCGTGGCGCTGATATCGTCCGTTATCGTACCGAGGATGAC 201
DB 130 ATCGCACAGTTCAGACGCAACCGGTTGTGACATCGTCCGCGTTGCTGCCCAAGACTGT 189
QY 202 GCGGCAAGCGCTCAAACTCATCAAAACGAGGTTAACTGCGCTGGTGGTGCATCATC 261
DB 190 GATGGAAGCACTGCCGATCATCGCAAGAGTCTCCGATCCAGTATCGCAGATATC 249
QY 262 CACTTCGACTATCGGATGCGCTGAAAGTACGCGGATACGCGGTGATTTGTCTGCGTATT 321
DB 250 CACTTCCAGCCCAAGTACATCTTCGCGCAATCGATGCAAGTTGCGCGCGCTTCGTGTG 309

QY 919 GTGAATGCCAGGTAGGCGCTGGTTTCTACACTCGCGCTCAACGGCGCAACAAGAA 978
Db 315193 GTTAATGGAATAGTGAATCTAAATAGCAACTTTAGGCTAGCAGGAAGTCATAAGAA 315252
QY 979 AGCGGCTCTATGAAGATGCGT---CGCAAGACCGTCTGGACAACAGATATGATC 1035
Db 315253 AGTGCAATTTATGAAGCGGAGTAAGACAATAAAGAAATAAAGAAACGAAGAAATATA 315312
QY 1036 GATCAGCTGGAAGCACGCAATTCGTGCGAA 1065
Db 315313 GAAAAATGGAATTAATAATTCGAAAAAA 315342

RESULT 4
US-09-974-300-1692
; Sequence 1692, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; PRIOR FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1692
; LENGTH: 1083
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-1692

Query Match 26.2%; Score 293.4; DB 10; Length 1083;
Best Local Similarity 56.6%; Pred. No. 3.4e-89;
Matches 543; Conservative 0; Mismatches 416; Indels 0; Gaps 0;

QY 22 CAACGTAGAAAATCAACACGATATTTACGTTGGGAATGTGCGGATTTGGCGATGTGCTCCC 81
Db 21 CATCGTACAAAACCGGCTCCGTTAAAGTGGACCTTTAAACAATAGGCGCAATAACGAA 80
QY 82 ATCGCGTACAGTCCATGACCAATACCGTACACAGAGCTCAAGCAACGGTCAATCAA 141
Db 81 GTCGTCAATTCAAAGCATGACAACAACGAAACACATGACGTTGAAGCAACCGTCCCGAA 140
QY 142 ATCAAGGCGCTGGAACGCGTTGGCGTGATATCGTCGGTGATATCCGTACCGACGATGGAC 201
Db 141 ATCAACAGACTCGGAGACGAGATGTCAATCGTCGGCTGCCCTGTCTCTGATGAACGG 200
QY 202 GCGGAGAAAGCGTTCAAATCATCAACACAGCAGTTAAAGTGGCGCTGGTGGCTGACATC 261
Db 201 GCTGCGAGCCATTCACAGAGATCAAAAGCGATATCCATCCCTCTTGTCTGGATATT 260
QY 262 CACTTCGATATCGATTTGGCTGAAAGTAGCGGAATACGGCGTCGATTTGTCTGCGTATT 321
Db 261 CATTTCAACTATAAATTTGGCATTTAAAGCGATCGAAGGCGGAGCCGATATAAATCCGCATC 320
QY 322 AACCCCTGGCAATATCGGTAATGAAGCGTATTTCGCGATGTGGTGTGACTGTGCGCGCAT 381
Db 321 AATCCGGGTAACTATCGGCGCCCGCAAAAGTTGAAGCGTCTCAACGCGAAGAGAA 380
QY 382 AAAAAATTCGATCCGATTTGCGTTAAAGCCGCGATCGCTGGAAAAAGATCTGCAAGAA 441
Db 381 AAGGCGATTCGATCCGATCCGATCGCGTCAATGACAGGCTCTCTGGAATAACGAATCCCTTGA 440
QY 442 AAGTATGGGAACCGACCGCGAGGCGTTGCTGGAATCTGCGCATGCGGTCAATGTTGATCAT 501
Db 441 AAGTACGCGTATCCGACAGCAGCGCATGGTGAAGGCGCCCTCGACCAACATTAAAT 500

RESULT 5

US-10-156-761-2548
; Sequence 2548, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 2548
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1155)
US-10-156-761-2548

Query Match 23.7%; Score 265; DB 14; Length 1155;
Best Local Similarity 55.5%; Pred. No. 1.8e-79;
Matches 533; Conservative 0; Mismatches 425; Indels 3; Gaps 1;

QY 43 ATTTACGTTGGAAATGTCCGATTTGGCGATGTGCTCCCATCGCGTACAGTCCATGACC 102
Db 70 ATCCAGGTGCGAACCCGTCGCGCGTGGCGGAGACGACCGCTCTCGGTGAGTTCGATGAG 129
QY 103 AATACGCGTACGACAGAGCTCGAAGCAACGGTCAATCAATCAAGCGCGCTGGAACGCGTT 162

Db 61 CCgATTGGCGATGGTGTCCATCGCGCTACAGTCCATGACCAATACGGCTACGACAGAC 120
QY 121 CTCGAGCAGCGTCAATCAATCAAGCGCTGGAAGCGTTGGCGCTGATATCGTCCGT 180
Db 121 CTCGAGCAGCGTCAATCAATCAAGCGCTGGAAGCGTTGGCGCTGATATCGTCCGT 180
QY 181 GTATCCGTTACCGACGATGAGCGCGGAGAGCGTTCAAACTCATCAAAAGAGGTTAAAC 240
Db 181 GTATCCGTTACCGACGATGAGCGCGGAGAGCGTTCAAACTCATCAAAAGAGGTTAAAC 240
QY 241 GTGCCGCTGGTGGCTGACATCCATTCGACTATCGCATTCGCTGCGTGAAGTAGCGGAATAC 300
Db 241 GTGCCGCTGGTGGCTGACATCCATTCGACTATCGCATTCGCTGCGTGAAGTAGCGGAATAC 300
QY 301 GGCGTCGATTGCTGGCTGATTAACCTGGCAATATCGGTAATCAAGAGCGTATTCGCATG 360
Db 301 GGCGTCGATTGCTGGCTGATTAACCTGGCAATATCGGTAATCAAGAGCGTATTCGCATG 360
QY 361 GTGGTTGACTGTGCGCGCGATAAAAACATTCGATCCGTTATGGCGTTAAACGCGGATCG 420
Db 361 GTGGTTGACTGTGCGCGCGATAAAAACATTCGATCCGTTATGGCGTTAAACGCGGATCG 420
QY 421 CTGGAAAAAGATCTGCAAGAAAAAGTATGCGAAACCGACGCGGTTGCTGGAATCT 480
Db 421 CTGGAAAAAGATCTGCAAGAAAAAGTATGCGAAACCGACGCGGTTGCTGGAATCT 480
QY 481 GCCATCGCTCATGTTGATCATCTCGATCGCTGAACTTCGATCAGTCAAGTCAGCGTG 540
Db 481 GCCATCGCTCATGTTGATCATCTCGATCGCTGAACTTCGATCAGTCAAGTCAGCGTG 540
QY 541 AAAGCGCTGACGCTCTCTCGCTGTTGAGTCTTATCGTTGCTGGCAAAACAGATCGAT 600
Db 541 AAAGCGCTGACGCTCTCTCGCTGTTGAGTCTTATCGTTGCTGGCAAAACAGATCGAT 600
QY 601 CAGCCGTTGATCTGGGGATCACGAAAGCGGTGTGCGCGACGCGGCGAGTAAATCC 660
Db 601 CAGCCGTTGATCTGGGGATCACGAAAGCGGTGTGCGCGACGCGGCGAGTAAATCC 660
QY 661 GCATTTGTTAGGTTAGTCTGCTGCTGCTGAGGATCGGCGACGCTGGCGGTATCGCTG 720
Db 661 GCATTTGTTAGGTTAGTCTGCTGCTGCTGAGGATCGGCGACGCTGGCGGTATCGCTG 720
QY 721 GCGGCGATCCGCTCGAAGAGATCAAGTCCGTTTCGATATTTGAAATCGCTGCGTATC 780
Db 721 GCGGCGATCCGCTCGAAGAGATCAAGTCCGTTTCGATATTTGAAATCGCTGCGTATC 780
QY 781 CGTTCGCGAGGATCAACTTCATCGCTGCGGACCTGTTTCGCGTCAAGAAATTTGATGT 840
Db 781 CGTTCGCGAGGATCAACTTCATCGCTGCGGACCTGTTTCGCGTCAAGAAATTTGATGT 840
QY 841 ATCGGTACGGTTAACCGCTGGAGCAACGCTGGAAGATATCATCTCGATGGAGCTT 900
Db 841 ATCGGTACGGTTAACCGCTGGAGCAACGCTGGAAGATATCATCTCGATGGAGCTT 900
QY 901 TCGATTATCGCTGCTGGTGAATGCGCCAGGTGAGCGCTGTTTCTACACTCGCGCTC 960
Db 901 TCGATTATCGCTGCTGGTGAATGCGCCAGGTGAGCGCTGTTTCTACACTCGCGCTC 960
QY 961 ACCGCGGCAACAAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGACCGTCTGGAC 1020
Db 961 ACCGCGGCAACAAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGACCGTCTGGAC 1020
QY 1021 AACACGATATGATCCAGCTGGAGCAACGCTTCGTCGGAAGCCAGTCAGCTGGAC 1080
Db 1021 AACACGATATGATCCAGCTGGAGCAACGCTTCGTCGGAAGCCAGTCAGCTGGAC 1080
QY 1081 GAAGCGGCTGCAATTGACGTTTACGAGGTTGAAAAATAA 1119
Db 1081 GAAGCGGCTGCAATTGACGTTTACGAGGTTGAAAAATAA 1119

RESULT 2

US-10-329-960-1

; Sequence 1, Application US/10329960
; Publication No. US2003009277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Fragment
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 15:08:02 ; Search time 412 Seconds
(without alignments)
8876.507 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119

Sequence: 1 atgcataacaggctccaat.....ttcagcaggtgaaaaataa 1119

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match %	Score	Length	ID	Description
1	1119	100.0	1119	9	US-09-921-992-3
2	622	55.6	1830121	14	US-10-329-960-1
3	403.6	36.1	640681	10	US-09-790-988-1
4	293.4	26.2	1083	10	US-09-974-300-1692
5	265	23.7	1155	14	US-10-156-761-2548
6	265	23.7	9025608	14	US-10-156-761-1
7	260.6	23.3	1155	14	US-10-156-761-1636
8	243.8	21.8	1164	10	US-09-712-363-1110
9	239.4	21.4	1134	10	US-09-738-626-2205
10	239.4	21.4	3309400	10	US-09-738-626-1
11	211.8	18.9	1180	10	US-09-881-752A-195
12	145.6	13.0	706	14	US-10-066-543-1184
13	70.6	6.3	584	9	US-09-921-992-21
14	68.4	6.1	596	9	US-09-921-992-23
15	68.4	6.1	670	9	US-09-921-992-22
16	56.8	5.1	348	10	US-09-974-300-6102

ALIGNMENTS

RESULT 1

US-09-921-992-3
; Sequence 3, Application US/09921992
; Patent No. US20020069426A1
; GENERAL INFORMATION:
; APPLICANT: Boronat, Albert;
; APPLICANT: Campos, Narciso;
; APPLICANT: Rodriguez-Concepcion, Manuel;
; APPLICANT: Rohmer, Michel;
; APPLICANT: Seeman, Myriam;
; APPLICANT: Valentin, Henry E.;
; APPLICANT: Venkatesh, Tyamagondlu V.;
; APPLICANT: Venkatesh, Mylavaram;
; TITLE OF INVENTION: Methyl-D-Erythritol Phosphate Pathway Genes
; FILE REFERENCE: 16516.107/35-21(51897)US
; CURRENT APPLICATION NUMBER: US/09/921,992
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/223,483
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 3
; LENGTH: 1119
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1119)
US-09-921-992-3

Query Match 100.0%; Score 1119; DB 9; Length 1119;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGCATACCCAGGCTCCCAATTCACGTCAGAAATCAACACGTATTACCTGGGAATGCG	60
Db	1	ATGCATACCCAGGCTCCCAATTCACGTCAGAAATCAACACGTATTACCTGGGAATGCG	60
Qy	61	CCGATTGGCGATGGTGGTCCCGTACAGTCCCGTACATCCCAATACGCGGTACGACAGAC	120

THIS PAGE BLANK (USPTO)

Db 541 GCTCCGGCACGGTGA 555

Search completed: November 23, 2003, 15:55:33
Job time : 96 secs

	Query Match	35.9%	Score 401.4	DB 4	Length 1137
;	TYPE: DNA				
;	ORGANISM: Acinetobacter baumannii				
US-09-328-352-3780	Best Local Similarity 60.8%				
	Matches 65	Conservative 0	Mismatches 421	Indels 0	Gaps 0
Qy	11	AGGCTCCAATTCAACGTAGAAAAATCAACACAGCTATTATTCGTTGGGAATGTGCCGATGTCGC 70			
Db	29	AGAACCCAAATTTAAACGTCGACCAACACGTAAATCCGTTGTTGGTCTATGTCGGTG 88			
Qy	71	ATGCTGCTCCCATCGCGTCAGTCCATGACCAATACCGGTACGACAGACGTCGAGCAA 130			
Db	89	CGCATGACCTATTAGTGTGCAAAAGTATGACAAATACCGAAACTTGCAGATGTGTGACGAA 148			
Qy	131	CGGTCAATCAAAATCAAGSCGCTGGAACCGGTTGGCGCTGATCGTCCGTGTATCCGTGTC 190			
Db	149	CTGTGGCTCAGATTGACGGTTGTGTGATGACAGGTGACAGATATTATGCGTGTTCAGTGC 208			
Qy	191	CGCATGGACGCGCGAGAACGCTTCAAACTCATCAAAACAGCAGGTAAACGTGCCGTGG 250			
Db	209	CTTCTATGGAAGCTGCTGAAGCCTTTGGTGCAAATCCGTAAGCGTGTTCAGTTCATTAG 268			
Qy	251	TGGCTGACATCCATCTCGACTATCGCATTTGGCTGAAAGTAGCGGAATACGGCGTCGATT 310			
Db	269	TAGCTGATATCCATTTTGACCATAGAAATGGCTTTGGCAGTTGTCAGATATTATGGTCG 328			
Qy	311	GTCTGCGGTATTAAACCTTGGCAATATCGGTAAATGAAGACGGTATTCGCATGGTGGTTGACT 370			
Db	329	GCTTTCGATTTAAACCGGGCAATATCGGTTGAGACCAAGAAAGTTCGTAAGTGTGGCGTG 388			
Qy	371	GTGCGCGCATAAAAACAATCCGATCCGATATGGCGTTAACCGCGGATCGCTGGAAAAAG 430			
Db	389	CGGCACGTATACCGGTATTCTATGCGTATTGGCGTAAATGCGCGGTTCTCTAGAAAAAG 448			
Qy	431	ATCTGCAAGAAAGTATGGCGAAACGAGCGCGAGGCGTTGCTGGAATCTGCCATGCGTC 490			
Db	449	ATTTACAGAAAAAATATGGCAGGCTACAGGCGAGGCACTTCTTGAAGTCAGCTTTACGCT 508			
Qy	491	ATGTTGATCATCTCGATCGCTGAACCTTCGATCAGTTCAAAAGTCAGCGTGAAGCGTCTG 550			
Db	509	ATATTGATATTTAGACCGTCTTGACTTCATGAGTTTAAAGTCAGTGAAGCATCAA 568			
Qy	551	ACGCTCTCTCGCTGTGTGAGTCTTATCGTTTCTGGGCAAAACAGATCGATCAGCGGTTCG 610			
Db	569	ATGTGTTTTTAAACCATGATGCTTATCGTTTTACTCTCTCAACAAATGATAATCCATTAC 628			
Qy	611	ATCTGGGGATCACCGAACCGCGTGTGCGCCGACGGGGCAGTAAATCCGCCATTGGTT 670			
Db	629	ACCTTGGAGTGACTGAAGCTGGTATTTACCGTACAGGTACTGTGAAATCAGCGATGTGCT 688			
Qy	671	TAGTCTCTGCTGCTCTGAAGGCATCGCGACACGCTCGCGGTATCGCTGGCGGCGCATC 730			
Db	689	TTGCTGGGTATTGATGGAAGGCATTGGCGATACGATCGGTATTTCGCTTCTGCTGAAC 748			
Qy	731	CGGTGCAAGAGATCAAAAGTCGGTTTCGATATTTTGAATATCGCTGCGTATCCGTTTCGCGAG 790			
Db	749	CTGAAGATGAAATCAAGATCGGTTTTTGATATCTTAAAAATCGTTGGCCTACGTTCTAACG 808			
Qy	791	GGATCAACTTCATCGCCTGCCGACCTGTTGCGGTACAGGAATTCATGTTATCGGTACGG 850			
Db	809	GTAATCAACTTTATTGCTTGTCCAAGTTGTTCTCGCCAAAGAAATTTAAACGTGATTCAGGTGA 868			
Qy	851	TTAAACGGCTGGAGCAACGCTTGAAGATATCATCACTCCGATGACGTTTCGATTATCG 910			
Db	869	TGCAGGCTTTAGAGAGCGTTTAGAGATATCCGTACACCAATGACGCTTCCTCGTTATTG 928			
Qy	911	GCTCGGTGGTGAATGGCCACAGGTGAGGCGCTGTTTCTACACTCGGCGTCAACCGCGCGGA 970			
Db	929	GCTGTAAGTAAATGGCCAGGTGAAGCAAAAGACAGATATCGGGTTGTTGGGCTG 988			

QY 661 GCCATTGGTTTAGGTCTGCTGCTCTGAAGGATCGCGACACCGTGCCTGATCGCTG 720
|||
Db 1467 GCGTGGGGTGGGATGCTCTTGGCGAGGAAATCGCGACACCATCCGGATTCCTTG 1408
|||
QY 721 GCGGCGATCCGTCGAGAGATCAAAAGTCGGTTTGCATATTTTGAATCGTCGATC 780
|||
Db 1407 GCTGCCGATCCGTCGAGGAGATCAAGTCGGTTTGCATATTCCTCAAGTCCTGCACCTG 1348
|||
QY 781 CFTTCGAGGGATCAACTTCATCGCTGCCGACCTGTTCCGTCAGGATTTGATGTT 840
|||
Db 1347 CGTCCCGTGGCATCAACTTCATCGCTGCCGACCTGTTTCGGCGAGAACTTCGACGTG 1288
|||
QY 841 ATCGGTACGGTTAAGCGCTGGAGCAACGCTCGAAGATATCATCTCCGATGGACGTT 900
|||
Db 1287 GTGAAGACCATGAACGAGCTGGAGGGCGCTGGAGACCTGCTGGTCGATGGACGTG 1228
|||
QY 901 TCATTTATCGGTCGCTGGTGAATGCGCCAGGTGAGCGCTGTTTCTACACTCGCGCTC 960
|||
Db 1227 GCGTGTATCGGTTGCGCTCGTCAACGGTCCGGCGAAGCCAAAGAGGCCATGTCCGCTC 1168
|||
QY 961 ACCGGCGGCAACGAAGACCGGCTCTATGAAGATGGCGTGGCAAGACCGTCTGGAC 1020
|||
Db 1167 ACCGGCGGCACTCCGAA---CCTGGTGTATATCGACGGCAAGCGCTGCGAAGAACTGACC 1111
|||
QY 1021 AACCAAGATATGATCGACCAAGCTGGAAGCAACGATTCGTGCGAAAGCCAGTCAAGCTGGAC 1080
|||
Db 1110 AAGCAACACTGTTGTCAGAGCTGGAAACGCGTATCGCGCAGAAAGGGCGCGAAGAGGCC 1051
|||
QY 1081 GAAGC 1085
|||
Db 1050 GAGGC 1046
|||

RESULT 13

US-09-252-991A-6339
; Sequence 6339, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6339
; LENGTH: 2088
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6339

Query Match 45.8%; Score 512; DB 4; Length 2088;
Best Local Similarity 68.0%; Pred. No. 1.8e-151;
Matches 729; Conservative 0; Mismatches 340; Indels 3; Gaps 1;

QY 14 CTCCAATTCAACGTAGAAAATCAACACGTATTTACCGTGGGAATGTGCGGATTTGGCGATG 73
|||
Db 6 CTCGGATCATTCGCCCAAGTCTCGAAATCTGGGTGCGCAACGTCCCGTGGCGCGCG 65
|||
QY 74 GTGCTCCCATCGCGTACAGTCATGACCAATACCGGTACGACAGACGTGCAAGCAACGG 133
|||
Db 66 ACGCGCGATCGCGTGCAGAGATGACCAACACCGAGACCTGCGACGTGCGTCCACCG 125
|||
QY 134 TCAATCAATCAAGCGCTGGAACGGTTGGCGCTGATATCGCTGATTCGTTACCGA 193
|||
Db 126 TCGCGAGATCGCGCCCTCGGAAGATGCGCGCGGACATCGTGGCGGTTTCGTTCCCG 185
|||
QY 194 CGATGACGCGGAGAAAGCGTTCAAACTCATCAACAGCAGGTTAAACGTGCGCGTGG 253
|||

Db 186 ACATGACGCGCGAGGCAATTCGCAAGATCAAGCAGAGGTCAACGTGCGCTGGTCG 245
|||
QY 254 CTGACATTCACATTCGACTATCGCATTCGCTGAAAGTAGGGATACGGCTGATGTGTC 313
|||
Db 246 CCGACATTCACATTCGACTATCGCATTCGCTGCGCTGCGCGAGTGGGAGTGGATGCC 305
|||
QY 314 TCGGTATTAACTCGCAATATCGGTAAATGAAGAGGTTATTCGATGCTGGTTGACTGTG 373
|||
Db 306 TGGGCATCAATTCGGGCAACATTCGTCGGAGGACCGGTCAGGCGGTGGTTCGATCCG 365
|||
QY 374 CCGCGGATAAAAACAATTCGATTCGATTCGCTGAAAGTAAACCGCGATCGCTGGAAAAAGATC 433
|||
Db 366 CCGCGAGCGCAACATTCGATTCGATTCGCTGAAAGTAAACCGCGATCGCTGGAAAAAGATC 425
|||
QY 434 TGCAGAAAGATATGCGCAACCGACCGCGAGGCGTTCGTAATCTGCCATCGCTCATG 493
|||
Db 426 TGCAGAAAGATATGCGCAACCGACCGCGAGGCGTTCGTAATCTGCCATCGCTCATG 485
|||
QY 494 TTGATCATCTCGATCGCTGAACTTCGATTCGATTCGCTGAAAGTAAACCGCGATCGCTGACG 553
|||
Db 486 TGCATCATCTCGAAGCTTGGACTTCAGAACTTCAGGTTCAGCTCAAGGCTCCGACG 545
|||
QY 554 TCTTCCTCGCTGTGAGTCTTATCGTTTCTGCGCAAAACAGATCGATCAGCGCTTCATC 613
|||
Db 546 TCTTCATGCGCTGCGCGCTATCGCTGCTGCGCAGGAGATCGAGAGCCCTCGAC 605
|||
QY 614 TGGGATACCAAGCGGTGTCGCGCAGCGGCGAGTAAATCCGCGATTTGGTTTAG 673
|||
Db 606 TGGGCATCACCGAGGCGCGCGCTGCGCTCGGCAACGCTGAAGTCCGCGTGGGCTGG 665
|||
QY 674 GTCTGCTGTCTGAAGGCATCGCGCACGCTGCGCTATCGCTGCGCGCGCGATCCGG 733
|||
Db 666 GCATGCTCTGCGCGAGGAATCGCGCACCATTCGGAATTCCTTCGCTGCGGATCCGG 725
|||
QY 734 TCGAAGAGATCAAGTTCGTTTCGATATTTTGAATTCGCTGCGTATTCCTTCGCGAGGA 793
|||
Db 726 TCGAGGAGATCAAGTTCGTTTCGATATTCGATATTCGATATTCGCTGCGCTCCCGTGGCA 785
|||
QY 794 TCACTTCATCGCTGCGCGCTGTCGCTGCGTCAAGATTTGATGTTTACGTTACGGTTA 853
|||
Db 786 TCACTTCATCGCTGCGCGCTGTCGCGCAGAACTTCGACGCTGTTGAAGACCATGA 845
|||
QY 854 ACGCGTGGAGCAACGCTGGAAGATATCATCACTCCGATGACGCTTCGATTCGCT 913
|||
Db 846 ACGAGCTGGAAGGCGCTTGGAGGACCTGCTGCTGCGATGAGCGTGGCGCTGATCGGTT 905
|||
QY 914 GCGTGTGAATGCGCCAGGTGAGGCGCTGTTTCTACACTCGCGCTGACCGCGGCAACA 973
|||
Db 906 GCGTGTCAACGCTCGGCGGAAGCCAGGAGGCCCATGTTCGCTCACCGCGGCACTC 965
|||
QY 974 AGAAAAGCGCTCTATGAAGATGCGTGGCGCAAGACCGTCTGGACAAACAGATATGA 1033
|||
Db 966 CGAA---CCTGGTGTATTCGACGCGCAAGCCCTGCGAGAAACTGACCAACGCAACCTGG 1022
|||
QY 1034 TCGACCACTGGAAGCAGCATTCGTTGCGAAAGCCAGTTCAGCTGCGACGAGC 1085
|||
Db 1023 TGGACAGCTGGAACGCTGATCCCGCAGAGGCGCGCGAGAAAGCCGAGGC 1074
|||

RESULT 14

US-09-328-3780
; Sequence 3780, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Berton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3780

Db 166 TTGCTTTGCCAACCTGTTACGCCAAGAAATTTGATGTGATTGTCACGGTAAATGCTTTGG 107
QY 863 AGCAAGCCCTGGAAGATATCATCACTCCGATGACGTTTCGATTATCGGCTGGTGTGA 922
Db 106 AGCAGCGCTCGAAGATATATCACGCCGATGATGTCCTATATTGTTGTGTAGTGA 47
QY 923 ATGGCCAGGTAGGCGCTGTTTCTACACTCGCGCTCACCGCGC 967
Db 46 ATGGCCCGGTGAAGCCGAGGTTTCTACTTACTTAGTGTGGCTGGCG 2

RESULT 10
US-09-170-187-10/c
; Sequence 10, Application US/09170187
; Patent No. 6381745
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medien & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,187
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/827,190
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; LENGTH: 886 base pairs
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-170-187-10

Query Match 47.8%; Score 534.6; DB 4; Length 886;
Best Local Similarity 75.3%; Pred. No. 8.2e-159;
Matches 666; Conservative 0; Mismatches 219; Indels 0; Gaps 0;
QY 83 TGGCCGTACAGTCCATGACCAATACCGTACGACAGACGTCGAAGCAACGGTCAATCAA 142
Db 886 TTGCTGTCCAATCTATGACGAATACGCGCACGCGGATGTTAGACCACTGTGCGGCAA 827
QY 143 TCAAGCGCTGGAACCGCTGGCGGTGATATCGTCGGTATCCGTACCGACGATGGACG 202
Db 826 TCCAATCACTTGAGCGGTGAGGTGTTGATATCGTCGCGGTGCTGTCTCAGCATGGATG 767
QY 203 CGGCAGACGGTTCAACTCATCAACACAGCGTTAACTGCGGTGCGGTGATCCGATCC 262
Db 766 CAGCAGACGCTTTAAATTAATTAACGACGCGGTGAATGTGCATTTGGTTGGGATATTC 707
QY 263 ACTTCGACTATCGCAITGGCGCTGAAAGTAGCGGAATACGCGCGTTCGATTGCTCGCGTATTA 322

Db 706 ACTTTGACTACCGTATCGCGATGAAAGTGGCTGAATATGTTGACTCCCTACGAATTA 647
QY 323 ACCCTGGCAATATCGTAAATGAAGAGCGTATTTCGATGTTGTTGACTCTGCGCGCGATA 382
Db 646 ACCCAGGTTAATATCGCAGTGAAGAGCGTATTTCGCCAAGTCGTTGATGATCTCGTCATC 587
QY 383 AAAACATTTCCGATCCGTTATTTGGCGTTAAACGCCGATTCGCTGGAAAAAGATCTGCAAGAAA 442
Db 586 ACAACATTTCTATCCGTATAGGGTCAATGGCGGGTCACTGGAAAAAGATATCCAAAGAA 527
QY 443 AGTATGGGAACCGACCGCGAGGCGTTGCTCGGAATCTGCCATGCGTCATGTTGATCATC 502
Db 526 AATACGGTGAAGCAACACCTGAAGCATTTGGTTGAATCAGCAATGCGACATGTTGATATCT 467
QY 503 TCGATCGCCTGAACCTTCGATTCAGTTTCAAGTTCAGGTGAGGCTGAAAGGCTGACGCTTCCTCG 562
Db 466 TGGACAGGCTGAATTTTCGATCAGTTTCAAGGTGAGTGTAAAGCGTCGGAATGCTTTCTTG 407
QY 563 CTGTTGAGTCTTATCGTTTGTCTGGCAAAACAGATCGATCAGCGCTTTCGATCTGGGGATCA 622
Db 406 CCGTGGCTCTTATCGTTTATTTGGCGCAAAAATTTGATCAACCACTTCACCTCGGTATTA 347
QY 623 CCGAAGCCGCTGCTGCGCGACGCGGCGAGTAAATTCGCCATTCGCTTTCGATCTGCTGC 682
Db 346 CAGAAAGCGGTGGGCTCGTTCTGTTTCAAGTGAATCAGCAATTTGCTTGGTATGTTGT 287
QY 683 TGTCTGAAGGCATCGCGACACGCTGCGGTATCGCTGCGGCGCGATCCGTCGAGAGA 742
Db 286 TGGCTGAAGGTATCGCGCATACGTTACGTTATCTACTCGCGCAGATCCTGTTGAGGAAG 227
QY 743 TCAAGTGGTTCGATATTTTGAATCGTCGCTATCGTTTCGGGAGGATCAACTTCA 802
Db 226 TGAAGTGGTTCGATATTTTGAATCGTTACGATCGCTCAGCTGCGATCAACTTTA 167
QY 803 TCGCTGCGCGACCTGTTGCGCTCAGGAATTTGATGTTTATCGGTACGGTTAAACGCGCTCG 862
Db 166 TTGCTTGCCCAACCTGTTTACGCCAAGAAATTTGATGTTGATGTTGTTGTTGTTG 107
QY 863 AGCAAGCGCTGGAAGATATCATCACTCCGATGAGCGTTTCGATTATCGGCTCGGTGTA 922
Db 106 AGCAGCGCTCGAAGATATTTATCACGCCGATGGATGTCCTATTATTGTTGTTGTTGTA 47
QY 923 ATGGCCCGGTGAGCGCGTGGTTTCTACACTCGCGGTCAACCGCGC 967
Db 46 ATGGCCCGGTGAGCGCGGTTTCTACTTTAGTGTGGCTGGCG 2

RESULT 11
US-09-252-991A-6420
; Sequence 6420, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6420
; LENGTH: 1644
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6420

Query Match 46.1%; Score 515.4; DB 4; Length 1644;
Best Local Similarity 67.8%; Pred. No. 1.3e-152;
Matches 736; Conservative 0; Mismatches 346; Indels 3; Gaps 1;

304 GTATCAATCCTGGCAACATTTGGTCTGTGAAGATCGCGTCCGCTGTTGATTTGGCGC 363
377 GGATTAACATTCGATCCGATCCGATTTGGCGTTAAACCGCGATCGCTGGAAGATCTGC 436
364 GAGACAAAATATTCGATTCGATTTGGTGTAAATGACGCTTTTAGAAAAGATTTC 423
437 AAGAAAATATGGCGAACCGACCGCGCGGCTTGTGGAATCTGCCATCGCGTCATGTTG 496
424 AAGAAAATATGGCGAACCAACCGCAGAGCCTTTGTAGAAATCCGATTCGGTCATGTAG 483
497 ATCATCTCGATCCGCTGAATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 556
484 AAATTCATGATCGCTTAACTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 543
557 TCCTCGCTGTTAGTCTTATCGTTTCTGCGCAACAGATCGATCAGCGCTTGCATCTCG 616
544 TCTTAGCGGTTGAACTTATTCGTTTACTGCTTAAGCAATTAACAGCCTTTACATTTAG 603
617 GGATCACGAAGCGGTGGTGGCGCGAGCGGGCAGTAAATCCGCCATTTGGTTTAGTTC 676
604 GCATTACAGAAGCAGGTGGCGCAGCGCTGTTGCAGTAAATCTGCAGTGGGTTTAGGAA 663
677 TCGTCTGTCTGAAGCATCGCGACACCGTTCGCGGTATCGCTGGCGGCGGATCCGCTCG 736
664 TGTATTAGCTGAGGCAATTTGGCGATACACTACGCGTCTCTTTGGCGGCGATTCCTGTAG 723
737 AAGAGATCAAGTCGCTTTCGATTTTGAATTCGCTGCGTATCGTTCGCGAGGATCA 796
724 AGGAATCAAGTCGCTTTCGATTTTGAATTCGCTGCGTATCGTTCGCGAGGATCA 783
797 ACTTCATCGCTCCGCGACCTGTTTCGCTCAGGAATTTGATGTTATCGGTACGGTTAAGC 856
784 ACTTTATTCGCTCCCAACCTGTTCTGCCAAGAAATTTGATGTTAATCGGTACAGTAAATG 843
857 CGCTGAGCAACCGCTTGAAGATATCATCTCCGATGACGTTTCGATTTACGCTGCGTGC 916
844 CGTGAACAACCGCTTGAAGATATTTATACCAATGGATGATCTATTTATCGGTTGTG 903
917 TGGTGAATGCGCCAGGTGAGGCGCTGTTTCTACACTCGCGCTCAGCGGCAACAAGA 976
904 TAGTGAATGCTTGGCGAGGCTACTGCTCCGATCTCGCGCTAACCGGCGGTAAACAAA 963
977 AAAGCGCCTTATGAAGATGGCGTGGCG --- AAAGACCGCTTGGACAAACAGATATGA 1033
964 AAAGCGTTATTATCTTTCGCGAGAGCGCCAAAGAGCGTTTGTATACGAAGATATAG 1023
1034 TCACACAGCTGGAAGCAGCGATTCGTGCGAAGCC 1069
1024 TGAACCAATTAGAAGCAAAATTCGTGCGAAGTC 1058

RESULT 9

US-08-827-190-10/c
; Sequence 10, Application US/08827190
; Patent No. 5858367
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/827,190
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 886 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-827-190-10
Query Match 47.8%; Score 534.6; DB 2; Length 886;
Best Local Similarity 75.3%; Pred. No. 8.2e-159;
Matches 666; Conservative 0; Mismatches 219; Indels 0; Gaps 0;
QY 83 TCGCGTACAGTCCATGACCAATACGCTACGACAGAGTCGAGCAACGGTCAATCAAA 142
DB 886 TTGCTGTCCAATCTATGACGAATACCGCACGACGATGTTGAAGCCACTGTGCGCAAA 827
QY 143 TCAAGGCGCTGAAACGCTTGGCGTGATATCGTCCGTGATCCGTACCGACGATGCGACG 202
DB 826 TCCAATACCTTGGCGTGTAGGTGTGATATCGTCCGCGTGTCTGTTCTACATGATGATG 767
QY 203 CGCGAAGACGCTTCAAACTCATCAACACAGCAGGTTAAACGTGCGCTGGTGGCTGACATCC 262
DB 766 CAGCAGAGCCCTTTAAATTAATTAAGCAGCGCTGAATGTGCAATTTGTTGCGGATATTC 707
QY 263 ACTTCGATATCGCATTTGGCGTGAAGTAGCGGAATACGCGTCAATGTTCTGCGTATTTA 322
DB 706 ACTTTGACTACCGTATCGCATGAAAGTGGCTGAATATGTTGTTGACTCCCTACCAATTA 647
QY 323 ACCCTGGCAATATCGTATGAAGACGCTTATTCGATGTTGCTGCTGCTGCGCGGATA 382
DB 646 ACCAGGTAAATACGCGAGTGAAGACGCTTATTCGCAAGTCGTTGATGCTGCTGCTATC 587
QY 383 AAAACATTCGCTATTCGCTTATTCGCTTAAACCGCGATCGCTGGAAGAAAGATTCGAAGAAA 442
DB 586 ACAACATTCCTATCCGTATAGGGTCAATGGCGGGTCACTGGAAGAAAGATATCCAGAAA 527
QY 443 AGTATGGCAACCGACCGCGCGAGGCTGCTGGAATCTGCCATGCGTCAATGTTGATCATC 502
DB 526 AATACGCTGAGCCAAACACCTGAAGCATTTGGTTGAATCAGCAATGCGACATGTTGATATCT 467
QY 503 TCGATCGCTGAACCTTCGATTCAGTTCAGAGTCAAGGTGGAAGCGTCAAGCTCTTCCTCG 562
DB 466 TGGACAGGCTGAATTTTCGATTCAGTTCAGGTGTTTAAAGCGTGGATGCTTTCTTCTG 407
QY 563 CTGTTGAGTCTTATCGTTTTCGCGCAACACATCGATCAGCGCTTTCATCTGGGATCA 622
DB 406 CGCTCGGCTTATCGTTTATTTGGCGCAAAAATTTGATCAACCACTTCACCTCGGTATTA 347
QY 623 CCGAAGCGGCTGCTGCGCGACGCGGCGAGTAAATTCGCCATTTGGTTAGGTCTGCTGC 682
DB 346 CAGAAGCGGCTGGGCTCGTTCTGTTCAAGTGAATCAGCAATTTGGTCTTGGTATTTGT 287
QY 683 TGTCTGAAGGCATCGCGACACGCTGCGGTATTCGCTGCGCGCGATCCGTTGCGAGAGA 742
DB 286 TGGCTGAAGGTATCGCGATACGTTACGTTATCTCACTCGCGCGAGATCTCTGTTGAGAG 227
QY 743 TCAAGTCTGTTTCGATATTTTGAATCTCGCTGCTATCCGTTCCGCGAGGATCAATCTCA 802
DB 226 TGAAGTCTGTTTGTATTTTGAATCTTAAATCGTTACGGATCCGCTCACGTGSCATCACTT 167
QY 803 TCGCTGCGCGACCTGTTTCGCTGAGGAATTTGATGTTATCGGTACGGTTAACCGCTGG 862

Best Local Similarity		74.9%	Pred. No. 1.1e-185;
Matches	790;	Conservative	0; Mismatches 262; Indels 3; Gaps 1;
Qy	17	CAATTCAACGTAGAAAAATCAACACGCTATTTTACGTTGGGAATGTGCGGATTCGGCATGGTG	76
Db	4	CTATTACGCGTCGTGAATCGCAAAAATTTATGTGGGAATGTACCAATTTGGTGGGATG	63
Qy	77	CTCCATGCGCGTACAGTCCATGACCAATACGCGTACGACAGCGTGAAGCAACGGTCA	136
Db	64	CGCCTATTGCGGTGCAATCAATGACAAATCTCGCACCTGATGTGGAAGCGACAGTTG	123
Qy	137	ATCAAAATCAAGGCGCTGGAACGCGTTGGCGCTGATATCGTCCGTGTATCCGTACCGACGA	196
Db	124	CTCAAAATTAATCAATTAAGACGCTGTGGTGAGATATTTGTTCGTGTATCTGTTCCAA	183
Qy	197	TGGACGGGCAAGACGCTTCAAACTCATCAACACGAGGTTAAAGTGCCTGCTGGTGGCGT	256
Db	184	TGGATGCTGCGAAGCATTTAAACAAATTAACAAACAGTGAATGTTCCCGTCTGTACAG	243
Qy	257	ACATCCACTTCGACTATCGCAATTCGCGTGAAGTAGCGGAATACGGGTCGATTGTCTGC	316
Db	244	ATAATTCATTTCGACTATCGTATCGCTTAAAGTCGAGAATATGGAAGTGAATGTTTAC	303
Qy	317	GTATTAAACCTGGCAATATCGGTAATGAAGAGCGTATTCGCATCGTGGTGTGACTGTGGC	376
Db	304	GTATCAATTCCTGGCAACATTTGGTTCGTGAAGATCGCGTCCGTGCGCTGTTGATTGTGCGC	363
Qy	377	CGGATAAAACATTTCCGATTCGCTATTCGGCTTAAAGTCGCGGATCGCTCGAAGAGATCTGC	436
Db	364	GAGACAAAATATTTCCGATTCGTTATGCTGTTAAATGACAGGCTCTTTAGAAAAAGATTTC	423
Qy	437	AAGAAAAGTATGGCGAAACCGACGCGGAGCGTTGCTGGAATCTGCCATCGCATGCTATG	496
Db	424	AAGAAAATATGGCAACCAACGCGCTGTTAGTAATTCGCATTCGCATCTAGT	483
Qy	497	ATCATCTCGATCGCCTGAACTTCGATCAGTTCAAAGTCAGCGTGAAGCGTGTGACGCT	556
Db	484	AAATTCCTAGATCGCTTAACTTCGATCAGTTTAAAGTGAGCGTAAAAAGCCTCCGATGTAT	543
Qy	557	TCCTCGCTGTAGGCTTATCGTTTGTGTCGCAACAGATCGATCAGCGCTTGCATCTGG	616
Db	544	TCTTAGCGGTGAATCTTATCGTTTATCTGGCTAAAGCAATTAACAGCCTTTACATTTAG	603
Qy	617	GGATCACCGAAGCCGCTGTGCGCGCACGCGGGCAGTAAATCCGCATTTGGTTAGGTC	676
Db	604	GCATTAAGAAGCAGGTGGCGCAGCGCTGGTGCAGTAAATCTGCAGTGGTTTAGAA	663
Qy	677	TGCTGCTGTCTGAAGGCATCGCGACACGCTGCGCGTATCGCTGGCGCGCGATCCGGTCG	736
Db	664	TGTTATTAGCTGAGGCAATGGCGATACATACGCGCTCTCTTTGGCGCGAGATCCTGTAG	723
Qy	737	AAGAGATCAAGTCGGTTTCGATTTTGAATCGCTGCGTATCCGTTCCGAGGATCA	796
Db	724	AGGAAATCAAGTCGGTTTGTATTTTGAATCTTTACGGATTCGTTTCAAGAGGAATTA	783
Qy	797	ACTTCATCGCTGCGCCACCTGTTCGCGTCAGGAAATTTGATTTATCGGTACGGTTAACG	856
Db	784	ACTTTATGCTTGCCCAACCTGTCTCGCAAGAAATTTGATTAATCGGTACAGTAAATG	843
Qy	857	CGCTGGAGCAACGCTCGGAAGATATCATCTCCGATGGAAGCGTTTCGATTATCGGCTGCG	916
Db	844	CGCTAGAAACACGCTTGAAGATATTATACACCAATGGATGTATCTATTATCGTTGTG	903
Qy	917	TGTTGANTGCCAGGTGAGGCGCTGTTCTACACTCGGCGGTACCGGGCGCAACAGA	976
Db	904	TAGTGAATGTCCTGGCGAGGCACTCGCTCCGATCTCGCGTAAACGGGGGTAAACAAA	963
Qy	977	AAAGCGGCTCTATGAAGATGGCGTGGCG--AAAGACCGCTCTGGACAAACAGATATGA	1033
Db	964	AAAGCGGTTATTATCTTGAAGGAGAACGCCAAAGAGCGTTTGTATTAACGAAGATATAG	1023
Qy	1034	TCGACCAAGCTGGAGCAAGCATTCGTCGGAAGCC	1068

Db 1024 TGAACCAATTAGACGACAAATTCGTGCGAAAGTC 1058

RESULT 8

US-09-170-187-7

Sequence 7, Application US/09170187

Patent No. 6383745

GENERAL INFORMATION:

APPLICANT: Rather, Philip N.

TITLE OF INVENTION: Methods For Screening For Antimicrobials

TITLE OF INVENTION: Utilizing aarC And Compositions Thereof

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/170,187

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/827,190

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Carroll, Peter G.

REGISTRATION NUMBER: 32,837

REFERENCE/DOCKET NUMBER: CASE-02443

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 1079 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

US-09-170-187-7

Query Match 55.4%; Score 619.8; DB 4; Length 1079

Best Local Similarity 74.9%; Pred. No. 1.1e-185;

Matches 790; Conservative 0; Mismatches 262; Indels 3

Qy 17 CAATTCAACGTAGAAAAATCAACACGCTATTTCAGTTGGGGAATGCGCGATGCGATTGGG

Db 4 CTATTAGCGTCGTGAATCGACAAAATTTATGTGGGAAATGTACCAATTGGT

Qy 77 CTCCTCATCGCGTACAGTCCATGACCAATACGCTACGACAGACGTCGGAAGC

Db 64 CGCTATTGCGTGCATCAATGACAAATCTCGCACCACTGATGTGGAAGC

Qy 137 ATCAAAATCAAGCGCTGGAACGCGTTGGCGCTGATATCGTCCGTGTATCCGTTAT

Db 124 CTCAAATTAATCATTAGACGCTGTTGGTGCAGATATGTTCTGTTATCTGTT

Qy 197 TGGACGGGCGAAGCGTTCAAACTCATCAAAACAGCAGGTTAACGTGCGCGCT

Db 184 TGGATGTCGCGAAGCATTTAAACAAATTTAAACAAACAGTGAATGTTCCGCT

Qy 257 ACATCCACTTCGATTCGCAATTCGGCTGAAAGTAGCGGAATACGCGCTCA

Db 244 ATATTCAATTCGATTCGATTCGGTGTAAAGTTCGCAAAATGAGTGAAGTGA

Qy 317 GTATTAAACCCCTGGCAATATCGGTAAATGAAGCGGTATTCGCATGGTGGTTGA

OPERATING SYSTEM:	MS DOS v6.22
SOFTWARE: ASCII Text	
CURRENT APPLICATION DATA:	
APPLICATION NUMBER:	US/09/643,990A
FILING DATE:	23-Aug-2000
CLASSIFICATION:	<Unknown>
PRIOR APPLICATION DATA:	
APPLICATION NUMBER:	08/487,429
FILING DATE:	1995-06-07
APPLICATION NUMBER:	08/426,787
FILING DATE:	1995-04-21
ATTORNEY/AGENT INFORMATION:	
NAME:	Kenley K. Hoover
REGISTRATION NUMBER:	40,302
REFERENCE/DOCKET NUMBER:	PB186PIC1
TELECOMMUNICATION INFORMATION:	
TELEPHONE:	301-610-5790
TELEFAX:	310-309-8439
INFORMATION FOR SEQ ID NO: 1:	
SEQUENCE CHARACTERISTICS:	
LENGTH:	1830121 base pairs
TYPE:	nucleic acid
STRANDEDNESS:	double
TOPOLOGY:	linear
SEQUENCE DESCRIPTION:	SEQ ID NO: 1;
US-09-643-990A-1	
Query Match	55.6%; Score 622; DB 4; Length 1830121;
Best Local Similarity	74.8%; Pred. No. 1.4e-184;

RESULT 5

US-09-557-884-1

; Sequence 1, Application US/09557884

; Patent No. 6506581

; GENERAL INFORMATION:

; APPLICANT: Fleischmann et al.

; TITLE OF INVENTION: The Nucleotide sequence of

; the Haemophilus influenzae Rd Genome, Fragments

; Thereof, and Uses Thereof

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville

; STATE: MD

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3 1/2 inch diskette

; COMPUTER: Dell Pentium

; OPERATING SYSTEM: MS DOS v6.22

; SOFTWARE: ASCII Text

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/557,884

; FILING DATE: 25-Apr-2000

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION NUMBER: 08/476,102

; FILING DATE: JUN-5-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Michelle S. Marks

; REGISTRATION NUMBER: 41,971

; REFERENCE/DOCKET NUMBER: PB186P3

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 301-309-8504

; TELEFAX: 301-309-8439

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1830121 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Query Match 55.6%; Score 622; DB 4; Length 1830121;

Best Local Similarity 74.8%; Pred. No. 1.4e-184;

Matches 794; Conservative 0; Mismatches 265; Indels 3; Gaps 1;

QY 10 CAGGCTCCCAATTCACAGTAGAATAATCAACACGATTTTACGTTGGGAATGTGCGGATTGGC 69
 DB 390972 CAGCCAACTATTNAGCGTGTGNAATCGACAAAATTTATGTGGAAATGTACCAATTGGT 391031

QY 70 GATGGTGTCTCCCATCGCGGTACAGTCCATCACCACCAATACGCGTACGACAGACGTCGAAGCA 129
 DB 391032 GGGGATGCGCTATTTCGCGTGAATCAATCACAATACTCGCACCACTGATGTGAAGCG 391091

QY 130 ACGGTCAATCAATCAAGCGGTGGAACGCGTTGGCGCTGATATCGTCCGTGATCCGTA 189
 DB 391092 ACAGTTGTCTCAATTAATTAATCATTAGAACGTTGTGGTGCAGATATTGTCGTGATCTGTT 391151

QY 190 CCGACGATGAGCGCGCAGAGCGTTCAAACTCATCAACACAGGTTAACTGCGCGGTG 249
 DB 391152 CCAACAAATGATGCTCGGAGCATTTAAACAAATTAACAAAGTGAATGTTCGCGTC 391211

QY 250 GTGGCTGACATCCACTTCGACTATCGCTGAAAGTAGCGGAATACGCGCGTCGAT 309
 DB 391212 GTAGCAGATATTCAATTCGACTATCGTATCGCGTTAAAGTGCAGAAATGAGTGGAT 391271

QY 310 TGTCTCGGTATTAAACCTCGCAATATCGGTATGAAGACGGTATTTCGATGGTGGTAC 369
 DB 391272 TGTTTACGTATCAATCTCGCAACATTTGGTTCGTGAAGATCGGTCGTCGCGTGGTGTAT 391331

QY 370 TGTGCGCGGATAAAAAACATTCCGATCCGATTTGGCGTTAACCGCGGATCGCTGGAAAAA 429
 DB 391332 TGTGCGCGAGACAAAAATATTCCGATTGCTATTGGTGTAAATGACGGCTCTTTTAGAAAAA 391391

QY 430 GATCTGCAAGAAAAAGTATGCGGAACCGACCGCGAGCGGCTTGGTGAATCTGCGCATCGCT 489
 DB 391392 GATTTTCAAGAAAAATATGCGGAACCAACCGCAGAGCCCTTGTAGAAATCCGCAATTGCGT 391451

QY 490 CATGTTGATCATCTCGATCGCTGAACTTCGATTCAGTTCAAAAGTACAGCTGAGAAAGCGTCT 549
 DB 391452 CATGTAGAAATTTCTAGATCGCTTAACTTCGATCAGTTTAAAGTGAGCGTAAAGCGCTCC 391511

QY 550 GACGCTTCTCCTCGCTGTTGAGTCTTATCGTTTGGCGAAAAACAGATCGATCAGCGCGTTG 609
 DB 391512 GATGTATTCTTAGCGGTTGAATCTTATCGTTTACTGGCTAAAGCAATTAACAGCGCTTTA 391571

QY 610 CATCTGGGATCACCGAAGCCGTTGGTGGCGAGCGGGGCGAGTAAATCCCGCATTTGTT 669
 DB 391572 CATTTAGGCATTTACAGAAGCAGGTGGCGCACGCGTGGTGCAGTAAATCTGCACTGGGT 391631

QY 670 TTAGTCTGCTGCTGCTCTGAAGGCATCGCGCACAGCTGCGCGTATCGCTGGCGCGGAT 729
 DB 391632 TTAGGAATGTTATTAGCTGAGGCGATTGGCGATACACTACGCGTCTCTTTGGCGCGAGT 391691

QY 730 CCGGTCGAAGAGATCAAAAGTCGGTTTCGATATTTTGAATTCGTCGATATCGTTTCGCGA 789
 DB 391692 CTTGTAGAGGAATCAAGTCGGTTTGNATATTTTGAATCTTTACGGATTTGTTCAAGA 391751

QY 790 GGGATCAACTTCATCGCTGCCGACCTGTTCCGCTCAGAGAAATTTGATGTTATCGGTACG 849
 DB 391752 GGAATTAACCTTTATTGCTTGGCCCAACCTGTTCTCGCAAGAAATTTGATGTAATCGGTACA 391811

QY 850 GTTAACGCGCTGGAGCAAGCCCTGGAAGATATCATCATCTCCGATCGACGTTTCGATTATC 909
 DB 391812 GTAAATGCGCTAGAACACGCGCTTGAAGATATTTATACCAATGGATGATCTATTATTC 391871

QY 910 GGCTCGTGTGTAATGGCCAGGTGAGGCGCTGGTTCCTACACTCGGCGTCACCGCGCGC 969
 DB 391872 GGTGTGTAGTAATGGTCTCGCGAGGCACTCGTCTCGATCTCGGCGTAACGGCGGT 391931

QY 970 AACAGAAAAAGCGGCTCTATGAAGATGGCGTGGCGC---AAAGCGCTCTGGACAAAC 1026
 DB 391932 AACAAAAAAAGCGGTTATTATCTTACGAGAGAACGCCAAAAAGAGCGTTTGTATAACGAA 391991

QY 1027 GATATGATCGACCACTGGAAGCAGCATTCGTGGGAAGCC 1068
 DB 391992 GATATAGTGAACCAATTAGAACCAAAAAATTCGTGCGAAAGTC 392033

RESULT 6

US-09-643-990A-1

; Sequence 1, Application US/09643990A

; Patent No. 6528289

; GENERAL INFORMATION:

; APPLICANT: Robert D. Fleischmann

; Mark D. Adams

; Owen White

; Hamilton O. Smith

; J. Craig Venter

; TITLE OF INVENTION: The Nucleotide sequence of
 ; the Haemophilus influenzae Rd Genome, Fragments
 ; Thereof, and Uses Thereof

; NUMBER OF SEQUENCES: 1

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Human Genome Sciences, Inc.

; STREET: 9410 Key West Avenue

; CITY: Rockville,

; STATE: MD

; COUNTRY: USA

; ZIP: 20850

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3 1/2 inch diskette

; COMPUTER: Dell Pentium

Db 1075 TCTATTATTGGTTGTGTAGTGAATGCCCGGTGAAGCCGAGGTTTCTACTTTAGGTGTG 1134
QY 961 ACCGGCGGCAACAGAAAACCGCCCTCTATGAAGATGGGTGGCC---AAAGACCGTCTG 1017
Db 1135 GCTGGCGCGAAAACCAAAAGTGGTTTCTATGAAGATGGCGTTTCGCAAAAAGAGCGTTT 1194
QY 1018 GACAAACAGATATGATCGACAGCTGGAAGCAGCAGATTCGTGCGAAAGCCAGTCAGCTG 1077
Db 1195 GATATGACATATTAATGATCAGCTTGAGCGCAAAATTCGCGCAAAAGCAGCAATGCTT 1254
QY 1078 GACGAA 1083
Db 1255 GATGAA 1260

RESULT 4

US-09-170-187-4
; Sequence 4, Application US/09170187
; Patent No. 6383745
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,187
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/827,190
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID. NO: 4:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 175..1272
; OTHER INFORMATION: /gene= "aarC"
; OTHER INFORMATION: /note= "Similar to E. coli GcpE protein listed by GenBank."
US-09-170-187-4

Query Match 56.4%; Score 631.6; DB 4; Length 1345;
Best Local Similarity 74.5%; Pred. No. 2.4e-189;
Matches 809; Conservative 0; Mismatches 274; Indels 3; Gaps 1;
QY 1 ATGCATAACAGCGCTCCAAATTCACGTAAGAAATCAACAGTATTTACGTTGGGAATGTG 60
Db 175 ATGCATAATGAATCACCGATAAAAGACGTAATCCACCGCAAAATTTATGTAGGTAACGTG 234

QY 61 CCGATTGGCGATGTGCTCCCATCCCGTACAGTCCATGACCAATACCGGTACGACAGAC 120
Db 235 CCTATTGGCGATGTGCTCCCATTCGTCCTCAATCTATGACGAATACCGCGCAGCGGAT 294
QY 121 GTCGAAGCAACCGGTCAATCAAAATCAAGGCGCTGGAACCGGTTGGCGCTGATATCTCGCT 180
Db 295 GTTGAAGCCACTGTGCGGCAAAATCCAATCACTTGAGCGGTGTAGGTGTTGATATCGTCCGC 354
QY 181 GTATCCGTACCGACCATGAGCGCGCAGAAAGGTTCAAATCTCATCAAAACAGCAGGTTAAC 240
Db 355 GTGTCTGTTCTTACCATGATGATGACAGAGAGCCTTTAAATTAATTAAGCAGCGCGTGAAT 414
QY 241 GTCCCGCTGGTGGCTGACATCCACTTCGACTATCGACTATGCGCTGAAAGTACGCGAATAC 300
Db 415 GTGCCATTGGTTGCGGATATTTCATTTGACTACCGTATCGCGATGAAAGTGGCTGAATAT 474
QY 301 GCGCTCGATTGTCTGCGGTATTAACCTTGGCAATATCGGTAAATGAAGAGCGTATTCGCTG 360
Db 475 GGTGTTGACTGCTGCTAGCAATTAACCCAGGTAATATCGGCAGTGAAGAGCGTATTCGCCAA 534
QY 361 GTGGTTGACTGCTGCGCGGATAAACATTCGATCCGTTATTTGGGCTTAACCCCGGATCG 420
Db 535 GTCGTTGATGAGTGTGCTGCTCATCAACATTCCTATCCGTATAGGGTCAATGCGCGGTCA 594
QY 421 CTGGAAAAGATCTGCAAGAAAAGTATGGCGAACCGACCGCCGACGCGTTGCTGGAATCT 480
Db 595 CTGGAAAAGATATCCAGAAAATACCGGTGAGCCACACACCTGAAGCATTTGGTTGAATCA 654
QY 481 GCATGCGTCTGTTGATCATCTCGATCGCTGAACTTCGATCAGTTCAAGTACGCGTG 540
Db 655 GCAATGCGACATGTTGATATCTTGGACAGCGCTGAATTTGATCAGTTCAAGGTCACTGTT 714
QY 541 AAAGCGCTGACGCTCTTCTCGCTGTGAGTCTTATGTTCTGTCGCAAAAACAGATCGAT 600
Db 715 AAAGCGTCGGATGTCTTCTCGCGTGGCTCTTATCGTTTATTTGGCGCAAAAATTTGAT 774
QY 601 CAGCGCTTGCATCTGGGGATCACCGAAGCGGTGGTGGCGCAGCGGGGACGTAATAATCC 660
Db 775 CAACCACTTCACTCGGTATTACAGAGCGGTGGGCTCGTTCTGGTTCACTGAATCA 834
QY 661 GCATTTGTTAGGTCTGCTGTCTGTAAGCGATCGCGCACACGCTGCGCGTATCGCTG 720
Db 835 GCAATTTGGTCTTGGTATGTTGTTGGCTGAAGGTATCGCGGATACGTTACGTATCTCACTC 894
QY 721 GCGGCGGATCCGTCGAAGAGATCAAGTCGCTTTGATATTTTGAATCGCTGCGGTATC 780
Db 895 GCGGCAGATCTCTGTTGAGGAAGTGAAGTCCGTTTGTATTTCTAAATTCGTTACCGATC 954
QY 781 CGTTCCGAGGATCAACTTTCATCGCTCCCGACCTGTTTCGCGTCAGGAATTTGATGTT 840
Db 955 CGCTCAGTGGCATCACTTTATGCTTGCCCAACCTGTTTACGCCAAGAAATTTGATGTG 1014
QY 841 ATCGGTACCGTTAACCGCGCTGAGCAACCGCTTGAAGATATCATCTCCGATGGAAGT 900
Db 1015 ATTGGTACGTTAAATGCTTTTGGAGCAGCGCTCGAAGATATTATCAGCGCGATGATGTC 1074
QY 901 TCGATTATCGCTGCGTGGTGAATGCGCCAGGTGAGCGCTGTTTCTACACTCGCGGTC 960
Db 1075 TCTATTATTGGTTGTGTAGTGAATGGCCGGGTGAAGCCGAGGTTTCTACTTTAGGTGTG 1134
QY 961 ACCGGCGCAACAAAGAAACCGCCCTCTATGAAGATGGGTGGCG---AAAGACCGTCTG 1017
Db 1135 GCTGGCGCAAAACCAAAAGTGGTTTCTATGAAGATGGCGTTTCGCAAAAAGAGCGTTT 1194
QY 1018 GACAAACAGATATGATCGACAGCTGGAAGCAGCAGTATTCGTGCGAAAGCCAGTCAGCTG 1077
Db 1195 GATAATGACAAATATTATTGATCAGCTTGAGCGCAAAATTCGCGCAAAAGCAGCAATGCTT 1254
QY 1078 GACGAA 1083
Db 1255 GATGAA 1260

Db 481 GCCATCGCTGATGTTGATCATCTCGATCGGCTGAACCTTCGATCACTGTTCAAAGTCAGCGTG 540
Qy 541 AAAGCGTCTACGCTCTCTCGCTGTTGAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600
Db 541 AAAGCGTCTACGCTCTCTCGCTGTTGAGTCTTATCGTTTGTGGCAAAACAGATCGAT 600
Qy 601 CAGCGCTGATCTGGGGATCACCGAAGCCGGTGGTGGCGGACGCGGGCAGTAAATCC 660
Db 601 CAGCGCTGATCTGGGGATCACCGAAGCCGGTGGTGGCGGACGCGGGCAGTAAATCC 660
Qy 661 GCCATTGGTTAGTCTGCTGCTGCTGAAGGATCGGCGACACGCTGCGCGTATCGCTG 720
Db 661 GCCATTGGTTAGTCTGCTGCTGCTGAAGGATCGGCGACACGCTGCGCGTATCGCTG 720
Qy 721 GCGGCCGATCCGGTCCGAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTCGCTATC 780
Db 721 GCGGCCGATCCGGTCCGAGAGATCAAAAGTCGGTTTCGATATTTTGAATCGCTCGCTATC 780
Qy 781 CGTTCGCGAGGGATCAACTTCATCGCTGCCCGACCTGTTTCGGGTCAAGAAATTTGATGTT 840
Db 781 CGTTCGCGAGGGATCAACTTCATCGCTGCCCGACCTGTTTCGGGTCAAGAAATTTGATGTT 840
Qy 841 ATCGGTACGTTTACCGGCTGGAGCAACGCTGCGGATATCATCTCCGATGGACGTT 900
Db 841 ATCGGTACGTTTACCGGCTGGAGCAACGCTGCGGATATCATCTCCGATGGACGTT 900
Qy 901 TCATTATCGCTGCGTGGTGAATGGCCGAGGCTGTTTCGAGATGGCGTGGCGAAAGA 1010
Db 901 TCATTATCGCTGCGTGGTGAATGGCCGAGGCTGTTTCGAGATGGCGTGGCGAAAGA 1010

RESULT 3

US-08-827-190-4
; Sequence 4, Application US/08827190
; Patent No. 5858367
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/827,190
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1345 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 175..1272
; OTHER INFORMATION: /gene= "aarC"
; OTHER INFORMATION: /note= "Similar to E. coli GopE protein listed by GenBank."
; US-08-827-190-4
Query Match 56.4%; Score 631.6; DB 2; Length 1345;
Best Local Similarity 74.5%; Pred. No. 2.4e-189;
Matches 809; Conservative 0; Mismatches 274; Indels 3; Gaps 1;
Qy 1 ATCATTAACACGAGCTCCAAATTCACCGTAGAATAACACGATATTTACGTTGGGAATGTG 60
Db 175 ATGCATAATGAATCACCGATATAAAGACGTAATCCACCGAATTTATGTAGTAACTG 234
Qy 61 CGATTGGCGATGGTCTCCCATCGCGGTACAGTCCATGACCAATACGGGTACGAGAC 120
Db 235 CCTATTGGCGATGGTCTCCCATTTGCTGTCCTCAATCTATGACGAATACGGGACGCGAT 294
Qy 121 GTCGAGCAACCGTCAATCAAAATCAAGGCGCTGGAACGGTTGGCGCTGATATCGTCGT 180
Db 295 GTTGAAGCACCTGTGGCGCAAAATCCAAATCACTGAGCGTGTAGGTGTTGATATCGTCCG 354
Qy 181 GTATCGGTACCGACGATGACGCGGAGAAAGCGTTCAAACTCATCAAAACAGAGGTTAAC 240
Db 355 GTGCTGTTCTTACGATGATGACGAGAAAGCGTTTAAATTAATTAAGCAGCGGTGAAT 414
Qy 241 GTGCGCTCGTGGGTGACATCCACTTCGACTATCGCATTCGCGTGAAGTACGGAATAC 300
Db 415 GTGCCATTGGTTGCGGATATTCATTTGACTACCGTATCGCATGAAAGTGGCTGAATAT 474
Qy 301 GCGTCGATGTCGCTGATTAACCTCGCAATATCGGTAAATGAAGAGCGTATTCGCATG 360
Db 475 GGTGTGACTGCTACGAAATTAACCCAGGTAATATCGGAGTGAAGAGCGTATTCGCCAA 534
Qy 361 GTGGTTGACTGTCGCGCGCATATAAACATTCGATCCGTTATGGCGTTTAAACGCGGATCG 420
Db 535 GTGCTTGTAGTGTCTGCTGCTCATCAACATCTCTATCCGTATAGGGGTCAATGGCGGTCA 594
Qy 421 CTGGAATAAGATCTCGAAGAAAGTATGCGAAACCGACCGCGAGCGGTTGTCTGGAATCT 480
Db 595 CTGGAATAAGATATCCAGAAATAACGCGTGAGCAACACCTGAAGCATTTGGTTGAATCA 654
Qy 481 GCCATGCGTCAATGTCATCTCGATCCCTGAACTTCGATCAGTTCAAAGTCAGCGTG 540
Db 655 GCAATCGCATGTTGATATCTTGGACAGGCTGAATTTTCGATCAGTTCAAGGTCACTGTT 714
Qy 541 AAAGGCTCTGACGCTCTTCCTCGCTGTTGAGTCTTATCGTTTCTGGCAAAACAGATCGAT 600
Db 715 AAAGGCTCGATGTTCTTCCTCGCTCTTATCGCTCTTATCGTTTATTTGGCGCAAAATTTGAT 774
Qy 601 CAGCGCTTGCATCTGGGGATCACCGAAGCCGGTGGTGGCGGACGCGGGCAGTAAATCC 660
Db 775 CAACCATTCACCTCGGTATTACAGAGCGGTGGGGCTCGTCTCGTTTCAAGTAAATCA 834
Qy 661 GCCATTGGTTAGTCTGCTGCTGCTGAAGGATCGGCGACACGCTGCGCGGTATCGCTG 720
Db 835 GCAATTTGGTCTTGGTATGTTGTTGGCTGAAGGATTCGGCGATACGTTACGTATCTCACTC 894
Qy 721 GCGGCCGATCCGTCGAAGAGATCAAGTCCGGTTTCGATATTTTGAATCGCTGCGGTATC 780
Db 895 GCGGCAGATCCCTGTTGAGGAAGTGAAGTCCGGTTTGTATTTTGAATTCGTTTACCGGATC 954
Qy 781 CGTTCGCGAGGATCAACTTCATCGCCTGCCGACCTGTTTCGCGTCAGGAATTTGATGTT 840
Db 955 CGCTCAGCTGGCATCAACTTTATTGCTGCCCAACCTGTTTACGCGCAAGAAATTTGATG 1014
Qy 841 ATCGGTAGGTTAAACGCGCTGGAGCAACGCTCGGAAGATATCATCTCCGATGACGTT 900
Db 1015 ATTGGTACGGTAAATGCTTTGGAGCAGCGCTCGAAGATATTTATCACGCGGATGATGC 1074
Qy 901 TCGATTATCGGCTGCGTGTGTGAATGGCCAGGTCAGGCGCTGTTTCTACACTCGGCGTC 960


```
Db 61 CCGATTGGCGATGGTCTCCATCGCGTACGTACCAATACCGGTACGACAGAC 120
Qy 121 GTGGAAGCAAGGTCAATCAATCAAGCGCTGGAACGGCTGGCGTGTATATCGCCGT 180
Db 121 GTGGAAGCAAGGTCAATCAATCAAGCGCTGGAACGGCTGGCGTGTATATCGCCGT 180
Qy 181 GTATCCGTACCGACGAGTGGAGCGGCGGAGAGCGTTCAAACTCATCAACAGCAGGTTAAC 240
Db 181 GTATCCGTACCGACGAGTGGAGCGGCGGAGAGCGTTCAAACTCATCAACAGCAGGTTAAC 240
Qy 241 GTCCCGCTGGTGGTGAATCCACTTCGACTATCGCATTCGCTGAAAGTAGCGGAATAC 300
Db 241 GTCCCGCTGGTGGTGAATCCACTTCGACTATCGCATTCGCTGAAAGTAGCGGAATAC 300
Qy 301 GGGCTGATTTGCTGGGTATTAACCCCTGGCAATATCGGTAAATGAAGAGCGTATTCGATG 360
Db 301 GGGCTGATTTGCTGGGTATTAACCCCTGGCAATATCGGTAAATGAAGAGCGTATTCGATG 360
Qy 361 GTGTTGACTGTCGGCGGAGTAAACATTCGATCCGATTCGCTTAAACGCGGATCG 420
Db 361 GTGTTGACTGTCGGCGGAGTAAACATTCGATCCGATTCGCTTAAACGCGGATCG 420
Qy 421 CTGGAAGCAAGATCTGCAAGAAAGTATGGCAACCGACGCGGAGGCTTGTGGAATCT 480
Db 421 CTGGAAGCAAGATCTGCAAGAAAGTATGGCAACCGACGCGGAGGCTTGTGGAATCT 480
Qy 481 GCCATGGTCAATGTCATATCGATTCGATTCGCTGAAAGTAGCGGAATAC 540
Db 481 GCCATGGTCAATGTCATATCGATTCGATTCGCTGAAAGTAGCGGAATAC 540
Qy 541 AAAGCGTCTGACGCTTCTCCGCTGTTGAGTCTTATCGTTGCTGCAAAACAGATCGAT 600
Db 541 AAAGCGTCTGACGCTTCTCCGCTGTTGAGTCTTATCGTTGCTGCAAAACAGATCGAT 600
Qy 601 CAGCCGTTGATCTGGGGATCACCGAAGCGGTTGTCGCGACGCGGCGAGTAAATCC 660
Db 601 CAGCCGTTGATCTGGGGATCACCGAAGCGGTTGTCGCGACGCGGCGAGTAAATCC 660
Qy 661 GCATTTGTTAGTCTGCTGCTGCTGAAGGATCGGCGACAGCTGCGCTATCGCTG 720
Db 661 GCATTTGTTAGTCTGCTGCTGCTGAAGGATCGGCGACAGCTGCGCTATCGCTG 720
Qy 721 GCGGCGATCCGCTGCAAGAGATCAAGTCCGTTTGCATATTTGAAATCGCTGCTATC 780
Db 721 GCGGCGATCCGCTGCAAGAGATCAAGTCCGTTTGCATATTTGAAATCGCTGCTATC 780
Qy 781 CGTTCGAGGAGTCAACTTTCATCGCTGCCGACCTGTTTCGCGTCAGGAATTTGATGT 840
Db 781 CGTTCGAGGAGTCAACTTTCATCGCTGCCGACCTGTTTCGCGTCAGGAATTTGATGT 840
Qy 841 ATCGGTACGGTTAACCGCTGAGAGCAACGCTCGAAGATATCATCTCCGATGAGCTT 900
Db 841 ATCGGTACGGTTAACCGCTGAGAGCAACGCTCGAAGATATCATCTCCGATGAGCTT 900
Qy 901 TCGATTATCGGCTGCTGTTGAAATGCCCCAGGAGGCGCTGTTCTACCTCGGCTC 960
Db 901 TCGATTATCGGCTGCTGTTGAAATGCCCCAGGAGGCGCTGTTCTACCTCGGCTC 960
Qy 961 ACCGGCGCAACAAAGAAAGCGGCTCTATGAAGATGGCGTGGCGCAAGA 1010
Db 961 ACCGGCGCAACAAAGAAAGCGGCTCTATGAAGATGGCGTGGCGCAAGA 1010
```

RESULT 2

```
US-09-170-187-8
; Sequence 8, Application US/09170187
; Patent No. 6381745
; GENERAL INFORMATION:
; APPLICANT: Racher, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,187
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/827,190
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1010 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-09-170-187-8
```

```
Query Match 90.3%; Score 1010; DB 4; Length 1010;
Best Local Similarity 100.0%; Pred. No. 6.4e-309;
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGCATAACCGAGCTCCCAATTCACGTAGAAAATCAACACGATATTACGTTGGGAATGTG 60
Db 1 ATGCATAACCGAGCTCCCAATTCACGTAGAAAATCAACACGATATTACGTTGGGAATGTG 60
Qy 61 CCGATTGGCGATGGTGTCTCCATCGCGTACAGTCCATGACCAATACGCTAGCAGAC 120
Db 61 CCGATTGGCGATGGTGTCTCCATCGCGTACAGTCCATGACCAATACGCTAGCAGAC 120
Qy 121 GTCGAAGCAACGGTCAATCAAAATCAAGCGCTTGGAAACGCGTTCGCTGATATCGTCGT 180
Db 121 GTCGAAGCAACGGTCAATCAAAATCAAGCGCTTGGAAACGCGTTCGCTGATATCGTCGT 180
Qy 181 GTATCCGTACCGACGATGAGCGCGGAGAGCGTTCAAACTCATCAACAGCAGGTTAAC 240
Db 181 GTATCCGTACCGACGATGAGCGCGGAGAGCGTTCAAACTCATCAACAGCAGGTTAAC 240
Qy 241 GTGCGCGCTGGTGGCTGACATCCACTTCGACTATCGCTGCGCTGAAAGTAGCGGAATAC 300
Db 241 GTGCGCGCTGGTGGCTGACATCCACTTCGACTATCGCTGCGCTGAAAGTAGCGGAATAC 300
Qy 301 GGGCTGATTTGCTGGGTATTAACCCCTGGCAATATCGGTAAATGAAGAGCGTATTCGATG 360
Db 301 GGGCTGATTTGCTGGGTATTAACCCCTGGCAATATCGGTAAATGAAGAGCGTATTCGATG 360
Qy 361 GTGTTGACTGTCGGCGGAGTAAACATTCGATCCGATTCGCTTAAACGCGGATCG 420
Db 361 GTGTTGACTGTCGGCGGAGTAAACATTCGATCCGATTCGCTTAAACGCGGATCG 420
Qy 421 CTGGAAGCAAGATCTGCAAGAAAGTATGGCAACCGACGCGGAGGCTTGTGGAATCT 480
Db 421 CTGGAAGCAAGATCTGCAAGAAAGTATGGCAACCGACGCGGAGGCTTGTGGAATCT 480
Qy 481 GCGATCGCTGATTTGCTGGGTATTAACCCCTGGCAATATCGGTAAATGAAGAGCGTATTCGATG 540
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 13:55:31 ; Search time 83 Seconds
(without alignments)
5950.695 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119

Sequence: 1 atgcataacaggctccaat.....ttcagcaggttgaaaaataa 1119

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA:*
- 1: /cgn2_6/prodata/1/ina/5A COMB.seq:*
 - 2: /cgn2_6/prodata/1/ina/5B COMB.seq:*
 - 3: /cgn2_6/prodata/1/ina/6A COMB.seq:*
 - 4: /cgn2_6/prodata/1/ina/6B COMB.seq:*
 - 5: /cgn2_6/prodata/1/ina/PCUTS COMB.seq:*
 - 6: /cgn2_6/prodata/1/ina/backfileseq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1010	90.3	1010	2	US-08-827-190-8
2	1010	90.3	1010	4	US-09-170-187-8
3	631.6	56.4	1345	2	US-08-827-190-4
4	631.6	56.4	1345	4	US-09-170-187-4
5	622	55.6	1830121	4	US-09-557-884-1
6	622	55.6	1830121	4	US-09-643-990A-1
7	619.8	55.4	1079	2	US-08-827-190-7
8	619.8	55.4	1079	4	US-09-170-187-7
9	534.6	47.8	886	2	US-08-827-190-10
10	534.6	47.8	886	4	US-09-170-187-10
11	515.4	46.1	1644	4	US-09-252-991A-6420
12	515.4	46.1	1644	4	US-09-252-991A-6131
13	512	45.1	2202	4	US-09-252-991A-6339
14	401.4	35.9	1137	4	US-09-328-352-3780
15	299	26.7	555	4	US-09-252-991A-6251
16	269	24.0	886	2	US-08-827-190-9
17	269	24.0	886	4	US-09-170-187-9
18	243.8	21.8	4403765	3	US-09-103-840A-2
19	243.8	21.8	4411529	3	US-09-103-840A-1
20	228.8	20.4	543	4	US-09-252-991A-6209
21	55.4	5.0	1083	4	US-09-221-017B-309
22	43.8	3.9	1230025	4	US-09-198-452A-1
23	40.6	3.6	1206	4	US-09-153-599A-2
24	40.6	3.6	2299	4	US-09-153-599A-1
25	38.4	3.4	1665	4	US-09-221-017B-1017
26	37.8	3.4	435	4	US-09-252-991A-14212
27	37	3.3	408	4	US-09-252-991A-1295

28	37	3.3	1257	4	US-09-252-991A-1225	Sequence 1225, Ap
29	37	3.3	2280	4	US-09-252-991A-1270	Sequence 1270, Ap
30	37	3.3	2616	4	US-09-252-991A-1336	Sequence 1336, Ap
31	35.4	3.2	583	3	US-08-998-416-1089	Sequence 1089, Ap
32	35.4	3.2	737	3	US-08-998-416-1039	Sequence 1039, Ap
33	35	3.1	1383	3	US-09-400-742-7	Sequence 7, Appli
34	35	3.1	1383	3	US-08-618-651A-7	Sequence 6, Appli
35	35	3.1	1383	4	US-09-215-252-6	Sequence 121, App
36	34.8	3.1	3312	4	US-08-311-731A-121	Sequence 121, App
37	34.6	3.1	1461	4	US-09-252-991A-5372	Sequence 5372, Ap
38	34.4	3.1	1476	4	US-09-252-991A-13090	Sequence 13090, A
39	34.2	3.1	1251	4	US-09-252-991A-444	Sequence 444, App
40	34.2	3.1	1293	4	US-09-252-991A-12207	Sequence 12207, A
41	34.2	3.1	1338	4	US-09-252-991A-12135	Sequence 12135, A
42	34.2	3.1	1362	4	US-09-252-991A-417	Sequence 417, App
43	34.2	3.1	1859	4	US-09-252-991A-12388	Sequence 12388, A
44	34.2	3.1	2334	4	US-09-620-312D-766	Sequence 766, App
45	34	3.0	8910	3	US-08-779-764A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-827-190-8
; Sequence 8, Application US/08827190
; Patent No. 5858367
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarC And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medlen & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/827,190
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1010 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-827-190-8

Query Match 90.3%; Score 1010; DB 2; Length 1010;
Best Local Similarity 100.0%; Pred. No. 6.4e-309;
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	ATGCAATACCAAGCTCCCAATTCACGTAGAGAAATCAACAGTATTACGTTGGGAATGTG	60
Db	1	ATGCAATACCAAGCTCCCAATTCACGTAGAGAAATCAACAGTATTACGTTGGGAATGTG	60
Qy	61	CCGATTGGCGATGGTCTCCCATCGGTACAGTCCATGACCAATACGCGTACGACGAC	120

```
CC      compositions
CC      thereof      Location/Qualifiers
PH      key          1.1010
FT      source       /organism='Escherichia coli',
                        Location/Qualifiers
                        1.1010
FEATURES
source   /organism='unidentified'
          /mol_type='genomic DNA'
          /db_xref='taxon:32644'
BASE COUNT      238 a      253 c      287 g      232 t
ORIGIN
Query Match      90.3%; Score 1010; DB 6; Length 1010;
Best Local Similarity 100.0%; Pred. No. 7.7e-239;
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1 ATGCATACCAGGCTCCAAATTCACGTAGAAAATCAACAGTATTTACGTTGGGAATGTG 60
      1 ATGCATACCAGGCTCCAAATTCACGTAGAAAATCAACAGTATTTACGTTGGGAATGTG 60
QY      61 CCGATTGGCGATGCTGCTCCCATGCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
      61 CCGATTGGCGATGCTGCTCCCATGCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
QY      121 GTGGAACCAACGGTCATCAATCAAGGCGTGAACGCGTGGCGCTGATATGTCCT 180
      121 GTGGAACCAACGGTCATCAATCAAGGCGTGAACGCGTGGCGCTGATATGTCCT 180
QY      181 GTATCCGTACCGACGATGACGCGCGAGAAAGCGTTCAAACTCATCAACAGCAGGTTAAC 240
      181 GTATCCGTACCGACGATGACGCGCGAGAAAGCGTTCAAACTCATCAACAGCAGGTTAAC 240
QY      241 GTGCGCTGCTGCTGCTGATTCATCTTGACTATGCAATGGCGTGAAGTACGCGAATAC 300
      241 GTGCGCTGCTGCTGCTGATTCATCTTGACTATGCAATGGCGTGAAGTACGCGAATAC 300
QY      301 GGGCTGATTTGTCTGCGATTTAAACCTGGGCAATTCGGTAAATGAAGCGGATTCGCGATG 360
      301 GGGCTGATTTGTCTGCGATTTAAACCTGGGCAATTCGGTAAATGAAGCGGATTCGCGATG 360
QY      361 GTGCTGACTGTGCGCGCGATTAATAACATTCGATCCGTAATGGCGTTAAAGCGCGATTCG 420
      361 GTGCTGACTGTGCGCGCGATTAATAACATTCGATCCGTAATGGCGTTAAAGCGCGATTCG 420
QY      421 CTGGAATAAAGATCTGCAAGAAAGATAGCGACGACGCGCGAGCGCTGCTGGAATCT 480
      421 CTGGAATAAAGATCTGCAAGAAAGATAGCGACGACGCGCGAGCGCTGCTGGAATCT 480
QY      481 GCCATGGGTCTATGTTGATCATCTCGATCGCTGAACCTTCATCACTCAAAAGTCAGCGTG 540
      481 GCCATGGGTCTATGTTGATCATCTCGATCGCTGAACCTTCATCACTCAAAAGTCAGCGTG 540
QY      541 AAAGCGTCTGACGCTTCTCTGCTGCTGATGCTTATGCTGCTGCGCAAAAACAGATCAT 600
      541 AAAGCGTCTGACGCTTCTCTGCTGCTGATGCTTATGCTGCTGCGCAAAAACAGATCAT 600
QY      601 CAGCGTTGCACTCTGGGAGTACCCGAAAGCGGTGCGCGAGCGGCGAGTAAATTC 660
      601 CAGCGTTGCACTCTGGGAGTACCCGAAAGCGGTGCGCGAGCGGCGAGTAAATTC 660
QY      661 GCCATTGGTTAAGTCTGCTGCTGCTGATGAGGCGATCGGCGACAGCTGCGGTAATCGCTG 720
      661 GCCATTGGTTAAGTCTGCTGCTGCTGATGAGGCGATCGGCGACAGCTGCGGTAATCGCTG 720
QY      721 GCGGCCGATCCGCTGCAAGAGATCAAAAGTCGTTTCATATTTGAATCGCTGCTATC 780
      721 GCGGCCGATCCGCTGCAAGAGATCAAAAGTCGTTTCATATTTGAATCGCTGCTATC 780
QY      781 CGTTGCGAGGAGATCAACTGCTGCTGCGACCTGTTGCGGTCAAGAAATTTGATGTT 840
      781 CGTTGCGAGGAGATCAACTGCTGCTGCGACCTGTTGCGGTCAAGAAATTTGATGTT 840
```

```
QY      841 ATCGTACCGTTTAACGCGCTGAGACAAACCGCTGGAAGATATCATCTCGATGACGTT 900
      841 ATCGTACCGTTTAACGCGCTGAGACAAACCGCTGGAAGATATCATCTCGATGACGTT 900
QY      901 TCGATTATCGGCTGCGTGTGTAATGCGCCAGGAGGAGGCGCTGTTCTACACTCGGCGTC 960
      901 TCGATTATCGGCTGCGTGTGTAATGCGCCAGGAGGAGGCGCTGTTCTACACTCGGCGTC 960
QY      961 ACCGGCGCAACAAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGA 1010
      961 ACCGGCGCAACAAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGA 1010
Db      ACCGGCGCAACAAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGA 1010
```

Search completed: November 23, 2003, 15:07:56
Job time : 4356 secs

```

Db      601 CAGCCGTCATCTGGGGATCACCAGACCGGTGTCGCGACGCGGAGCATTAATCC 660
Qy      661 GCCATTGGTTAGGTCGTGCTGTCTGAAGAGCATCGGCGACAGCTGCGGTAATCGCTG 720
Db      661 GCCATTGGTTAGGTCGTGCTGTCTGAAGAGCATCGGCGACAGCTGCGGTAATCGCTG 720
Qy      721 GCGGCCGATCCGTCGAGAGATCAAAAGTCGGTTTCATATTTTGAATGCTGCGTATC 780
Db      721 GCGGCCGATCCGTCGAGAGATCAAAAGTCGGTTTCATATTTTGAATGCTGCGTATC 780
Qy      781 CGTTGCGAGAGGATCACTTCATGCGCTGCGCCGACCTGTTGCGCTCAGAGATTTGATGT 840
Db      781 CGTTGCGAGAGGATCACTTCATGCGCTGCGCCGACCTGTTGCGCTCAGAGATTTGATGT 840
Qy      841 ATCGGATCGGTTAAAGCGCTGAGAGCAAGCGCTGAGAGATATCATCATCCGATGAGCTT 900
Db      841 ATCGGATCGGTTAAAGCGCTGAGAGCAAGCGCTGAGAGATATCATCATCCGATGAGCTT 900
Qy      901 TCGATTATCGGCTGCTGCTGTAATGAGCCAGGTGAGGCGCTGTTTCTACACTCGGCGTC 960
Db      901 TCGATTATCGGCTGCTGCTGTAATGAGCCAGGTGAGGCGCTGTTTCTACACTCGGCGTC 960
Qy      961 ACCGCGCGCAACAAGAAAGCGGCTCTATGAGATGCGGTGCGCAAGA 1010
Db      961 ACCGCGCGCAACAAGAAAGCGGCTCTATGAGATGCGGTGCGCAAGA 1010

RESULT 14
LOCUS      AR208328 1010 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 8 from patent US 6383745.
ACCESSION  AR208328
VERSION     AR208328.1 GI:21509454
KEYWORDS
SOURCE      Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 1010)
            Rather, P.N.
TITLE       Methods of screening for anti-microbial utilizing aarc and
            compositions thereof
            Patent: US 6383745-A 8 07-MAY-2002;
            Location/Qualifiers
            source          1..1010
            /organism="unknown"
BASE COUNT  238 a 253 c 287 g 232 t

Query Match      90.3%; Score 1010; DB 6; Length 1010;
Best Local Similarity 100.0%; Pred. No. 7,7e-239;
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db      301 GCGGTGATGTGTGGGTATTAACCTGGCAATATCGTATGAGAGCGTATTCGATG 360
Qy      361 GTGGTTGACGTGCGCGGATTAATAACATTCGATCCGTAATGGCGTTAAAGCGCGATCG 420
Db      361 GTGGTTGACGTGCGCGGATTAATAACATTCGATCCGTAATGGCGTTAAAGCGCGATCG 420
Qy      421 CTGGAATAAGATTCGAAAGAAAGTATGCGAAGCAGCGCGCGGCTGCTGATCT 480
Db      421 CTGGAATAAGATTCGAAAGAAAGTATGCGAAGCAGCGCGCGGCTGCTGATCT 480
Qy      481 GCCATCGTCATGTTGATCATCTGATCGCTGAGACTTGCATGATTCATCAAGTCAGCGTG 540
Db      481 GCCATCGTCATGTTGATCATCTGATCGCTGAGACTTGCATGATTCATCAAGTCAGCGTG 540
Qy      541 AAGCGCTGACGCTCTTCTCGCTGTTGATGCTTATTCGTTTGTGCGCAAAACGATCGAT 600
Db      541 AAGCGCTGACGCTCTTCTCGCTGTTGATGCTTATTCGTTTGTGCGCAAAACGATCGAT 600
Qy      601 CAGCCGTCATCTGGGATACACCAAGCCGCTGAGCGGCGAGCGGGGAGTAAATCC 660
Db      601 CAGCCGTCATCTGGGATACACCAAGCCGCTGAGCGGCGAGCGGGGAGTAAATCC 660
Qy      661 GCCATTGGTTAGGTCGTGCTGTCTGTAAGGATCGGGAACACGCTGCGGCTATCGCTG 720
Db      661 GCCATTGGTTAGGTCGTGCTGTCTGTAAGGATCGGGAACACGCTGCGGCTATCGCTG 720
Qy      721 GCGGCCGATCCGTCGAGAGATCAAAAGTCGGTTTGAATATTCGCTGCTATC 780
Db      721 GCGGCCGATCCGTCGAGAGATCAAAAGTCGGTTTGAATATTCGCTGCTATC 780
Qy      781 CGTTGCGAGAGGATCACTTCATGCGCTGCGGACCTGTTGCGGTGAGGAATTTGATGT 840
Db      781 CGTTGCGAGAGGATCACTTCATGCGCTGCGGACCTGTTGCGGTGAGGAATTTGATGT 840
Qy      841 ATCGGATCGGTTAAAGCGCTGAGAGCAAGCGCTGAGAGATATCATCATCCGATGAGCTT 900
Db      841 ATCGGATCGGTTAAAGCGCTGAGAGCAAGCGCTGAGAGATATCATCATCCGATGAGCTT 900
Qy      901 TCGATTATCGGCTGCTGCTGTAATGAGCCAGGTGAGGCGCTGTTTCTACACTCGGCGTC 960
Db      901 TCGATTATCGGCTGCTGCTGTAATGAGCCAGGTGAGGCGCTGTTTCTACACTCGGCGTC 960
Qy      961 ACCGCGCGCAACAAGAAAGCGGCTCTATGAGATGCGGTGCGCAAGA 1010
Db      961 ACCGCGCGCAACAAGAAAGCGGCTCTATGAGATGCGGTGCGCAAGA 1010

RESULT 15
LOCUS      BD084908 1010 bp DNA linear PAT 27-AUG-2002
DEFINITION Methods for screening for antimicrobials utilizing (aarc) and
            compositions thereof.
ACCESSION  BD084908
VERSION     BD084908.1 GI:22630518
KEYWORDS
SOURCE      unidentified
            unidentified
            unclassified.
            unclassified.
REFERENCE   1 (bases 1 to 1010)
            Rather, P.N.
TITLE       Methods for screening for antimicrobials utilizing (aarc) and
            compositions thereof
            Patent: JP 2001523097-A 6 20-NOV-2001;
            CASE WESTERN RESERVE UNIVERSITY
            OS Escherichia coli
            PN JP 2001523097-A/6
            PD 20-NOV-2001
            PF 27-MAR-1998 JP 1998546016
            PR 27-MAR-1997 US 08/827190
            PI PHILIP N RATHER
            PC C1201/68, C07K14/24
            CC Methods for screening for antimicrobials utilizing (aarc) and

```

Matches 1103; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

```

QY 1 ATGATTAACAGGCTCCATTCAACGTAGAAAACACAGCTATTTCGTGGAAATG 60
DB 280346 ATGATTAACAGGCTCCATTCAACGTAGAAAACACAGCTATTTCGTGGAAATG 280287
QY 61 CCGATTGGCATGTGCTCCCATGCGCTACAGTGCATGACCAATACGCGTACAGAC 120
DB 280286 CCGATTGGCATGTGCTCCCATGCGCTACAGTGCATGACCAATACGCGTACAGAC 280227
QY 121 GTGAGAACAAAGGTCAATCAATCAAGCGCTGAGACGCGTGGCGCTGATTCCT 180
DB 280226 GTGAGAACAAAGGTCAATCAATCAAGCGCTGAGACGCGTGGCGCTGATTCCT 280167
QY 181 GTATCCGTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 280166 GTATCCGTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 280107
QY 241 GTGCGCGTGGTGGTGCATTCATTCATTCATTCATTCATTCATTCATTCATTC 300
DB 280106 GTGCGCGTGGTGGTGCATTCATTCATTCATTCATTCATTCATTCATTCATTC 280047
QY 301 GCGCGTATTGTCTGCTGATTAACCTTGGCAATATCGGTAATGAGAGCGTATTC 360
DB 280046 GCGCGTATTGTCTGCTGATTAACCTTGGCAATATCGGTAATGAGAGCGTATTC 279987
QY 361 GTGTTGATCTGTGGCGCGATTAACATTCCTGATTCCTGATTCCTGATTCCTG 420
DB 279986 GTGTTGATCTGTGGCGCGATTAACATTCCTGATTCCTGATTCCTGATTCCTG 279927
QY 421 CTGGAAGAAAGATCTGCAAGAAAGATATGCGAACCGACGCGAGCGTGTCTGATCT 480
DB 279926 CTGGAAGAAAGATCTGCAAGAAAGATATGCGAACCGACGCGAGCGTGTCTGATCT 279867
QY 481 GCCATGCGTATGTGATTCATTCCTGATTCCTGATTCCTGATTCCTGATTCCTG 540
DB 279866 GCCATGCGTATGTGATTCATTCCTGATTCCTGATTCCTGATTCCTGATTCCTG 279807
QY 541 AAGCGCTACGCTTCCTGCTGCTGATTCCTGATTCCTGATTCCTGATTCCTGATTC 600
DB 279806 AAGCGCTACGCTTCCTGCTGCTGATTCCTGATTCCTGATTCCTGATTCCTGATTC 279747
QY 601 CACCGCTTGATCTGAGGATCACCGAAGCGGTGTCGCGACGCGGAGAGTAATTC 660
DB 279746 CACCGCTTGATCTGAGGATCACCGAAGCGGTGTCGCGACGCGGAGAGTAATTC 279687
QY 661 GCCATTGTTAGTGTCTGCTGCTGATTCCTGATTCCTGATTCCTGATTCCTGATTC 720
DB 279686 GCCATTGTTAGTGTCTGCTGCTGATTCCTGATTCCTGATTCCTGATTCCTGATTC 279627
QY 721 GCGGCGCATTCGCTGCGAAGATCAAGTGGTTGATTCCTGATTCCTGATTCCTGATTC 780
DB 279626 GCGGCGCATTCGCTGCGAAGATCAAGTGGTTGATTCCTGATTCCTGATTCCTGATTC 279567
QY 781 CGTTGCGAGAGATCACTTCATTCCTGCTGCGACGCTGCTGCGAATTCGATTC 840
DB 279566 CGTTGCGAGAGATCACTTCATTCCTGCTGCGACGCTGCTGCGAATTCGATTC 279507
QY 841 ATCGGTAAGGTTAAAGCGCTGAGCAAGCGCTGAGATATCATCTCCGATGAGCTT 900
DB 279506 ATCGGTAAGGTTAAAGCGCTGAGCAAGCGCTGAGATATCATCTCCGATGAGCTT 279447
QY 901 TCGATTATCGGCTGCGGTGATGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 960
DB 279446 TCGATTATCGGCTGCGGTGATGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 279387
QY 961 ACCGCGCGCAACAAGAAAGCGCTCTATGAGATGAGCGTGCAGAAAGCGCTGAGAC 1020
DB 279386 ACCGCGCGCAACAAGAAAGCGCTCTATGAGATGAGCGTGCAGAAAGCGCTGAGAC 279327
QY 1021 AACACGATATGATGACAGAGCTGAGAGACGATTCGTGCGAAGCGCACTCAGCTGAGAC 1080
DB 279326 AACACGATATGATGACAGAGCTGAGAGACGATTCGTGCGAAGCGCACTCAGCTGAGAC 279267

```

QY 1081 GAAGCGCTGCAATTGACGCTTCAAGGTTGAAAATTA 1119
DB 279266 GAAGCGCTGCAATTGACGCTTCAAGGTTGAAAATTA 279228

RESULT 13
AR027993 1010 bp DNA linear PAT 29-SEP-1999
LOCUS Sequence 8 from patent US 5858367.
ACCESSION AR027993
VERSION AR027993.1 GI:593966
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 1010)
AUTHORS Rather, P.N.
TITLE Methods for screening for antimicrobials utilizing Aac and compositions thereof
JOURNAL Patent: US 5858367-A 8 12-JUN-1999;
FEATURES Location/Qualifiers
source 1..1010 /organism="unknown"
BASE COUNT 238 a 253 c 287 g 232 t

Query Match 90.3%; Score 1010; DB 6; Length 1010;
Best Local Similarity 100.0%; Pred. No. 7.7e-239;
Matches 1010; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGATTAACAGGCTCCATTCAACGTAGAAAACACAGCTATTTCGTGGAAATG 60
DB 1 ATGATTAACAGGCTCCATTCAACGTAGAAAACACAGCTATTTCGTGGAAATG 60
QY 61 CCGATTGGCATGTGCTCCCATGCGCTACAGTGCATGACCAATACGCGTACAGAC 120
DB 61 CCGATTGGCATGTGCTCCCATGCGCTACAGTGCATGACCAATACGCGTACAGAC 120
QY 121 GTGAGAACAAAGGTCAATCAATCAAGCGCTGAGACGCGTGGCGCTGATTCCT 180
DB 121 GTGAGAACAAAGGTCAATCAATCAAGCGCTGAGACGCGTGGCGCTGATTCCT 180
QY 181 GTATCCGTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
DB 181 GTATCCGTACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 240
QY 241 GTGCGCGTGGTGGTGCATTCATTCATTCATTCATTCATTCATTCATTCATTC 300
DB 241 GTGCGCGTGGTGGTGCATTCATTCATTCATTCATTCATTCATTCATTCATTC 300
QY 301 GCGGCGCATTCGCTGCGAAGATCAAGTGGTTGATTCCTGATTCCTGATTCCTG 360
DB 301 GCGGCGCATTCGCTGCGAAGATCAAGTGGTTGATTCCTGATTCCTGATTCCTG 360
QY 361 GTGTTGATCTGTGGCGCGATTAACATTCCTGATTCCTGATTCCTGATTCCTG 420
DB 361 GTGTTGATCTGTGGCGCGATTAACATTCCTGATTCCTGATTCCTGATTCCTG 420
QY 421 CTGGAAGAAAGATCTGCAAGAAAGATATGCGAACCGACGCGAGCGTGTCTGATCT 480
DB 421 CTGGAAGAAAGATCTGCAAGAAAGATATGCGAACCGACGCGAGCGTGTCTGATCT 480
QY 481 GCCATGCGTATGTGATTCATTCCTGATTCCTGATTCCTGATTCCTGATTCCTG 540
DB 481 GCCATGCGTATGTGATTCATTCCTGATTCCTGATTCCTGATTCCTGATTCCTG 540
QY 541 AAGCGCTGAGCTTCCTGCTGCTGATTCCTGATTCCTGATTCCTGATTCCTGATTC 600
DB 541 AAGCGCTGAGCTTCCTGCTGCTGATTCCTGATTCCTGATTCCTGATTCCTGATTC 600
QY 601 CACCGCTTGATCTGAGGATCACCGAAGCGGTGTCGCGACGCGGAGAGTAATTC 660
DB 601 CACCGCTTGATCTGAGGATCACCGAAGCGGTGTCGCGACGCGGAGAGTAATTC 660

```

(hyb clm)

DGKGVKLDSDNIRLLFLSLAGNYLANONPDVHOAKLLEDESKIEFLVLSLDEMTBSA
KADVLLPETSFMERMINIETWGTASLILSEKPIEDFERDIDYDLRVAKLGVE
AFBSQDEKOWIEHIMEORLAMPDRIOTRELPALRKLRSKAPLAFENINDP
ONNPPRPSKIKIEFSKRLPDMODPEIPALSHVYPRAREGDEKLTAKYPIQLITWYK
NRANSTOYANPMIOEVOTKLMNPDAKRGISBEGSVKIVNDRGSIIPVEITPPI

Query Match 97.7%; Score 1093.4; DB 1; Length 11521;

Best Local Similarity 98.6%; Pred. No. 2e-259;

Matches 1103; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 ATGCATACCAAGGCTCCATTCAACGTAGAAAATCAACAGGATTTACGTTGGAGATGTG 60
DB 1204 ATGCATACCAAGGCTCCATTCAACGTAGAAAATCAACAGGATTTACGTTGGAGATGTG 1145
QY 61 CCGATTGGCGATGTGTCCCATGCGGTACAGCCATGACCAATAGCGGTACGACGAC 120
DB 1144 CCGATTGGCGATGTGTCCCATGCGGTACAGCCATGACCAATAGCGGTACGACGAC 1085
QY 121 GTGGAAGCAACGGTCAATCAAAATCAAGGCGGTGAGCGGTGCGGTATATCGTCGT 180
DB 1084 GTGGAAGCAACGGTCAATCAAAATCAAGGCGGTGAGCGGTGCGGTATATCGTCGT 1025
QY 181 GTATCCGTAACGAGATGAGCGCGGCAAGAGCTTCAAACTCATCAACAGCAGTTAAC 240
DB 1024 GTCTCCGTAACGAGATGAGCGCGGCAAGAGCTTCAAACTCATCAACAGCAGTTAAC 965
QY 241 GTGCGCGGTGAGCGGTCAATCCATTCATTCGCAATTCGCGTAAGTAAGCGGAATAC 300
DB 964 GTGCGCGGTGAGCGGTCAATCCATTCATTCGCAATTCGCGTAAGTAAGCGGAATAC 905
QY 301 GGGGTGATTTGCTGCTGATTTAAACCTGCGCAATTCGTAATGAAAGCGTATTCGATG 360
DB 904 GGGGTGATTTGCTGCTGATTTAAACCTGCGCAATTCGTAATGAAAGCGTATTCGATG 845
QY 361 GTGGTGAATGCTGCGCGCGATTAACCAATTCGATTCGTAATGAAAGCGTATTCGATG 420
DB 844 GTGGTGAATGCTGCGCGCGATTAACCAATTCGATTCGTAATGAAAGCGTATTCGATG 785
QY 421 CTGGAAGAAATGCGAAGAAAGTATAGCGAACCGGCGGCGGCTGCGGAATCT 480
DB 784 CTGGAAGAAATGCGAAGAAAGTATAGCGAACCGGCGGCGGCTGCGGAATCT 725
QY 481 GCCATGCGTATGTTGATCATCTCGATCGCTGAACTTCGATCGTTCAAAAGTCAGCGTG 540
DB 724 GCCATGCGTATGTTGATCATCTCGATCGCTGAACTTCGATCGTTCAAAAGTCAGCGTG 665
QY 541 AAAGCGTCTGACGCTCTTCTGCTGCTGTTGAGTCTTATCGTTGCTGGCAAAAACAGATGAT 600
DB 664 AAAGCGTCTGACGCTCTTCTGCTGCTGTTGAGTCTTATCGTTGCTGGCAAAAACAGATGAT 605
QY 601 CAGCGCTTGCATCTGGGGATCAACGAAAGCGGTGTGCGCGACGGGGCGAGTAAATCC 660
DB 604 CAGCGCTTGCATCTGGGGATCAACGAAAGCGGTGTGCGCGACGGGGCGAGTAAATCC 545
QY 661 GCCATTGTTAAGGTGCTGCTGCTGCTGAAAGCATCGGCGCAACGCTGCGGATTCGCTG 720
DB 544 GCCATTGTTAAGGTGCTGCTGCTGCTGAAAGCATCGGCGCAACGCTGCGGATTCGCTG 485
QY 721 GCGGCGCATCGGTGGAAGAGATCAAAAGTGGTTTCAATATTTTGAATCGTCCGATC 780
DB 484 GCGGCGCATCGGTGGAAGAGATCAAAAGTGGTTTCAATATTTTGAATCGTCCGATC 425
QY 781 CGTTCCGAGGAGATCAATCTTATGCTGCTGCGGACCTGTTGCGGTCAAGAAATTTGATGT 840
DB 424 CGTTCCGAGGAGATCAATCTTATGCTGCTGCGGACCTGTTGCGGTCAAGAAATTTGATGT 365
QY 841 ATCGGTACGTTAAACGCGTGAACAACGCTGGAAGATATCATCATCCGATGAGAGTT 900
DB 364 ATCGGTACGTTAAACGCGTGAACAACGCTGGAAGATATCATCATCCGATGAGAGTT 305
QY 901 TCGATTATCGGCTGCTGTTGAATGCGCCAGTGAAGCGCTGCTGTTTCAACTCGGCGTC 960
DB 304 TCGATTATCGGCTGCTGTTGAATGCGCCAGTGAAGCGCTGCTGTTTCAACTCGGCGTC 245

QY 961 ACCGGCGGCAACAAGAAAAGCGGCTCTATGAGATGCGGTGGCAAGACCGTCTGAC 1020
DB 244 ACCGGCGGCAACAAGAAAAGCGGCTCTATGAGATGCGGTGGCAAGACCGTCTGAC 185
QY 1021 AACACGATATGATTCAGCAGCAGTGAAGACCGGATTCGTGCGAAGCCAGTCAGTTGAC 1080
DB 184 AACACGATATGATTCAGCAGCAGTGAAGACCGGATTCGTGCGAAGCCAGTCAGTTGAC 125
QY 1081 GAAGCGCGTTCGATTTGACGTTCAAGCAGTTGAAATTA 1119
DB 124 GAAGCGCGTTCGATTTGACGTTCAAGCAGTTGAAATTA 86

RESULT 12
AP002561/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

1
Kurokawa, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S.,
Kurokawa, K., Ishii, K., Hattori, M., Tetsuno, I., Abe, H., Iida, T.,
Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T.,
Sasakawa, C., and Shingawa, H.
Complete nucleotide sequence of the prophage VT2-Sakai carrying the
verotoxin 2 genes of the enterohemorrhagic *Escherichia coli* O157:H7
derived from the Sakai outbreak
Genes Genet. Syst. 74 (5), 227-239 (1999)
2
Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M.,
Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shingawa, H. and
Hayashi, T.
Comparative analysis of the whole set of rRNA operons between an
enterohemorrhagic *Escherichia coli* O157:H7 strain MG1655
and *Escherichia coli* K-12 strain MG1655
Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
3
Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S.,
Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasakawa, C. and
Shingawa, H.
Complete nucleotide sequence of the prophage VT1-Sakai carrying the
Shiga toxin 1 genes of the enterohemorrhagic *Escherichia coli*
O157:H7 strain derived from the Sakai outbreak
Gene 258 (1-2), 127-139 (2000)
4
Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
Tanaka, M., Toke, T., Iida, T., Takami, H., Honda, T., Sasakawa, C.,
Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
Shingawa, H.
Complete genome sequence of enterohemorrhagic *Escherichia coli*
O157:H7 and genomic comparison with a laboratory strain K-12
DNA Res. 8 (1), 11-22 (2001)
5 (bases 1 to 296827)
Omishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shingawa, H. and
Hayashi, T.
Direct Submission
Submitted (126-JUN-2000) Ken Kurokawa, Osaka University, Genome


```

/translate_table=11
/product="orf, hypothetical protein"
/protein_id="AAG57625.1"
/db_xref="GI:12516909"
/translation="MNRQAPIQRKSRITRVVNPICDGPAPIAVOSMTTRTTDEAT
VNOIKALERVGADIVRSVPPTMDAAEFKLIKQOVNPLVADIHFDRIALKYAEVGV
DCIRINGNIGNEERIIMVVDICARDKNIPIRIGNAGSLERKDIQEKYEPTPLGVS
AMRHVHLDRINLDFQFVSKASDVPLAVESYRLAKOIDQPLHGLITTEAGARGAV
KSAIGLGLLSEGGDLRVSLAADPVEIKVGGDLKSLRISRGINFIACPTCSQ
EPVIGTVNALRQRLIEDIITPMOVSIICGVNGRGELVLTGLTGKNKSGLYEDCV
RKRLDNDMDIDQLEAIRAKASQLEBARLDVQVQR"
/gene="yfgA"
/complement(1231..2244)
/feature="putative membrane; Not classified"
/notes="Residues 1 to 337 of 337 are 99.10 pct identical to
residues 1 to 337 of 337 from Escherichia coli K-12 strain
MG1655; B2516"
/codon_start=1
/translate_table=11
/product="putative membrane protein"
/protein_id="AAG57626.1"
/db_xref="GI:12516910"
/translation="MNTTEATHDONALTTGARLRNAREQLGSGQAVAEELCKVSTV
RDLEEDAPADIASLTPRGYRSARLVHPEEELRGLERKAPLRARAKVAPQSSRL
GKRKKKDGMLMTFTLVLFVVGLSGAMWQHKAOEITTMADSSXELSSNSQ
GVSPLNTATTTTPATTSTPPASVDITATOTPVVYAPAPAVDPOQNAVSSQNSQ
DTATAPATTAATPDGAAPLPTDQAGVTTPVADPNMLVMNFTADCMLEVDANGKLI
FSGQRKGNLNLTGQAPYKLIKIGAPAAVAVQYOGKPVDSLRFTRTQVARTLNAEQS
PAQ"
/complement(2529..3683)
/gene="yfgB"
/feature="synonym: Z3780"
/complement(2529..3683)
/gene="yfgB"
/function="orf; Unknown function"
/notes="Residues 1 to 384 of 384 are 99.47 pct identical to
residues 1 to 384 of 384 from Escherichia coli K-12 strain
MG1655; B2517"
/codon_start=1
/translate_table=11
/product="orf, hypothetical protein"
/protein_id="AAG57627.1"
/db_xref="GI:12516911"
/translation="MSBQLVTPENVTTKQKINLLDLNRQMEFFKDLGEKPRADQ
VMKMYHYCCDNFEMTDINKVLRGKLEVAERABEVREORSDDGTTIKMAIYADQ
RVEVYIPEDDRATLCVSGQCALCEKFCSTAQGFNRRLRVSEIIQGVRAKIVG
AAVTGORPITVVMGMGEPLINNVNPAEMILDDGFGLSKRVTLSTGCVPA
LEPLGMDIVDALAISLHAPRDEIRDEIVPINKXKXNETPLAAVRYRLKSNANQGV
IEVYMLDHYVDGTEHQAHLKDLDPCKINLIPMFPFPAPIGRSSNSRIDFSSKVL
MSYFTTIVAKRTGGDDIDAACQGLADVDVDRTRTLRKMGGEAIDIKAV"
/complement(3833..4264)
/gene="ndk"
/feature="synonym: Z3781"
/complement(3833..4264)
/gene="ndk"
/function="enzyme; Nucleotide biosynthesis: Purine
ribonucleotide biosynthesis"
/notes="Residues 1 to 143 of 143 are 100.00 pct identical
to residues 1 to 143 of 143 from Escherichia coli K-12
Strain MG1655; B2518"
/codon_start=1
/translate_table=11
/product="nucleoside diphosphate kinase"
/protein_id="AAG57628.1"
/db_xref="GI:12516912"
/translation="MAIERFSLIKRPAVAKVINGNIPARFEAGFKYICTKMLHLY
EQARKYAEHDGKPFDPDGLVEFMTSGPIVSVLEGNAVQRRHDLGATNPANALAGT
LRADYADSLTENTHGSDSVESAAREIAVFFGGEVCPTR"
/feature="misc_feature"
/notes="O-island #106; Region of the EDL933 chromosome not
```

```

homologous to E. coli K-12 MG1655"
/complement(4405..4983)
/gene="Z3782"
/complement(4405..4983)
/gene="Z3782"
/function="orf; Unknown function"
/notes="No significant matches"
/codon_start=1
/translate_table=11
/protein_id="AAG57629.1"
/db_xref="GI:12516913"
/translation="MLMKRQGITAIADLSQAKAWERALLLNELCTROYEGEWSV
SPHTSEVNVSRRLFFVQADVOMTQIOGLRHLHLMPOISPARPMPOKOLCGA
CMRACQOVSLENGHQLINDALCNCQNCIACVFOAMTVELTILPAKIVLNAHRK
VCKTCKSFLTFQONQNCITCOGRHRYGNRTP"
/complement(5259..6080)
/gene="Z3783"
/complement(5259..6080)
/gene="Z3783"
/function="putative enzyme; Not classified"
/notes="Residues 2 to 272 of 273 are 25.26 pct identical to
residues 6 to 276 of 286 from Genpept 118 :
g1|5002126|gb|AAD37317.1|AF135170_8 (AF135170) dimethyl
sulfoxide reductase subunit C [Yersinia pestis]"
/codon_start=1
/translate_table=11
/product="putative dimethyl sulfoxide reductase subunit C"
/protein_id="AAG57630.1"
/db_xref="GI:12516914"
/translation="MHEPLPMAFTLLQASVGTWLMARLHRIILITSSARPYMRPL
LAFLAGAVGLISTHLGYPPLANMALSHFSSWLSREITFCALYALGLSTLVM
LGMGQMLLMALAAVGVDFVCMQOVYHTSIITWQHYNTLIMFGTILGSLALV
VRIISGILPQIDALRNGCVIYALYELIYQPLMGDLTANQOIRLPHRPLML
GQLPILTSKGIISYGMNPFVAVGGCKNIPALRFSVNLVGSSEVNLRFVFSIG"
/complement(6073..6510)
/gene="Z3784"
/complement(6073..6510)
/gene="Z3784"
/function="putative enzyme; Not classified"
/notes="Residues 4 to 134 of 145 are 61.83 pct identical to
residues 57 to 186 of 205 from Genpept 118 :
g1|1787122|gb|AAC73981.1| (AE000191) anaerobic dimethyl
sulfoxide reductase subunit B [Escherichia coli]"
/codon_start=1
/translate_table=11
/product="putative dimethyl sulfoxide reductase subunit B"
/protein_id="AAG57631.1"
/db_xref="GI:12516915"
/translation="MSHNVFAYTLVSCHNCADPICKPKPTPMAMHRRPGDGIIVNT
DKVCGCYACMACPYGAPQMNQOTGMSXCDICIDQAEOPICVATPELAKIRGP
IDLEBKRYDSDVDKGLPPSSITQPLVYIKPHQGEKESHHH"
/complement(6699..9080)
/gene="Z3785"
/complement(6699..9080)
/gene="Z3785"
/function="putative enzyme; Not classified"
/notes="Residues 25 to 792 of 793 are 43.59 pct identical
to residues 3 to 784 of 785 from Genpept 118 :
g1|1787121|gb|AAC73980.1| (AE000191) anaerobic dimethyl
sulfoxide reductase subunit A [Escherichia coli]"
/codon_start=1
/translate_table=11
/product="putative dimethyl sulfoxide reductase subunit A"
/protein_id="AAG57632.1"
/db_xref="GI:12516916"
/translation="MDTPBEKERITFPQVRSRSPLOATSLTLPLPISSTAKQSPDA
SPENVAPVADKVPNCSTFDDGCKDIDAHMGDGVATOTTLIDNLDPOMPIMRACV
RGRGRKPYHPHDLRIKPYKRGKGEKGFESISDEATTLIADNKRITTOYGRASR
YHVGAVAGGTFSGPAMARRLNLTGGLTYEYSVSGNTAAATYITGVASGSM
DTLLDNLKLVLMGNHPTETIFGHYHTFIDTYTLGFDENSMGPVANSLSLADQIP
LPLTPLEWAEKRTIVPAOSIROLARDVYTRPAALIQMGQORHICGRTARGSTLL
ASITQNVGIGKGMAAGYGSSSNRKFCVGPDMENPVQAKISIMNMQADADASKVTPQ
```



```
gene      complement (7286. .8551)
          /gene="yfcB"
          /locus_tag="c2876"
          complement (7286. .8551)
CDS
          /gene="yfcB"
          /locus_tag="c2876"
          /EC_number="2.1.1.72"
          /function="putative"
          /note="Escherichia coli K-12 ortholog: b2330; Escherichia
          coli O157:H7 ortholog: z3593"
          /codon_start=1
          /transl_table=1
          /product="Hypothetical adenine-specific methylase yfcB"
          /protein_id="AAN81326.1"
          /db_xref="GI:26109124"
          /translation="MAGIGFLNOLPWHFADPPTLRSDNRVILNLASPRHOLAKOR
          LVIAKCCFLHLSFSINPTPTSTTDREKRTPHSPRNADLSSSSQTSRRN
          YCHALBEEVDKIFVDEAVNELQTLQMLWVSFRFAANIWGHGDTNPDWDAVOLY
          LPSLYPLDIPEDMRTARLTSSEKRLVEVIRVRVERIIVAVLTNRAMPGEFYVD
          ERVLVPSPIGELINNKFAGLISKQPHIDMCTGSGCIAIACPAPEAEVADVIS
          PDALVAEONIEBGLIHNVIPIRSDLFRLPKVQYDLVTNPVYDAMEMDLPNEY
          RHPBELGASGTGKLTRLIRILGNADYLDADVILICEVNSVHLMEOYPDVPTWML

Query Match      98.1%; Score 1098.2; DB 1; Length 300099;
Best Local Similarity 98.8%; Pred. No. 1,4e-260;
Matches 1106; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy      1  ATGCATACAGAGGTCATTCACAGTAAGAAATCAACAGGTATTTACGTTGGAAATG 60
Db      188410 ATGCATACAGAGGTCATTCACAGTAAGAAATCAACAGGTATTTACGTTGGAAATG 188351

Qy      61  CCGATTGGCGATGTCCTCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
Db      188350 CCGATTGGCGATGTCCTCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 188291

Qy      121  GTCGAACCAACGGTCAATCAATCAAGCGCTGGAACGCGTGGCGCTGATTCGTCCT 180
Db      188290 GTCGAACCAACGGTCAATCAATCAAGCGCTGGAACGCGTGGCGCTGATTCGTCCT 188231

Qy      181  GTATCCGTCACCGACGATGACCGGCGGACGCGTCAACTCATCAACGACGATTAC 240
Db      188230 GTATCCGTCACCGACGATGACCGGCGGACGCGTCAACTCATCAACGACGATTAC 188171

Qy      241  GTCGCGGTGGTGGTGCATCCATCTTGATCGCATTCGCTGCGTGAAGAGCGGAATAC 300
Db      188170 GTCGCGGTGGTGGTGCATCCATCTTGATCGCATTCGCTGCGTGAAGAGCGGAATAC 188111

Qy      301  GGGGTGATGTCGCTGATTAACCTTGCAATATCGGTATGAAGACGCTATTCGATG 360
Db      188110 GGGGTGATGTCGCTGATTAACCTTGCAATATCGGTATGAAGACGCTATTCGATG 188051

Qy      361  GTGTGATCTGTGCGGCGCATTAACATTCGATCCGTATTTGGCGTTAACCGCGGATCG 420
Db      188050 GTGTGATCTGTGCGGCGCATTAACATTCGATCCGTATTTGGCGTTAACCGCGGATCG 187991

Qy      421  CTGGAATAAAGATCTGAAGAAGATAGCGAACCGACCGCGGAGGCGTTGCTGAATCT 480
Db      187990 TTGGAATAAAGATCTGAAGAAGATAGCGAACCGACCGCGGAGGCGTTGCTGAATCT 187931

Qy      481  GCCATGCGTATGTTGATCATCTCGATCGCCTGAACCTTCGATCAATTCAAGTCAGCTG 540
Db      187930 GCCATGCGTATGTTGATCATCTCGATCGCCTGAACCTTCGATCAAGTCAGCTG 187871

Qy      541  AAGCGTCTGAGCTCTTCTCGCTGTTGAGTCTTATTCGTTGCTGGCAAAACAGATCGAT 600
Db      187870 AAGCGTCTGAGCTCTTCTCGCTGTTGAGTCTTATTCGTTGCTGGCAAAACAGATCGAT 187811

Qy      601  CAGCGTGTGATCTGGGAGTCAACGAAACCGGTGCGCGGCGACCGGCGGACGTAATATCC 660
Db      187810 CAGCGTGTGATCTGGGAGTCAACGAAACCGGTGCGCGGCGACCGGCGGACGTAATATCC 187751

Qy      661  GCCATTGTTAGTCTGCTGCTGTAAGGATCGGCGACACGCTGCGGATCGCTG 720
Db      187750 GCCATTGTTAGTCTGCTGCTGTAAGGATCGGCGACACGCTGCGGATCGCTG 187691
```

```
Db      187750 GCTATTGTTTAACTGCTGCTGCTGTAAGGATCGGCGACACGCTGCGGATCGCTG 187691
Qy      721  GCGCGCATCCGTCGCAAGATCAAAAGCGTTTGATATTTGAATGCTGCGATC 780
Db      187690 GCGCGCATCCGTCGCAAGATCAAAAGCGTTTGATATTTGAATGCTGCGATC 187631

Qy      781  CGTTTCGAGGATCACTTCATCGCTGCCACCTGTTCCGTCGCAAGATTTGATGTT 840
Db      187630 CGTTTCGAGGATCACTTCATCGCTGCCACCTGTTCCGTCGCAAGATTTGATGTT 187571

Qy      841  ATCGTACGTTTACCGCTGAGCAACGCTGGAAGATATCATCTCCGATGACGTT 900
Db      187570 ATCGTACGTTTACCGCTGAGCAACGCTGGAAGATATCATCTCCGATGACGTT 187511

Qy      901  TCGATTATGCGCGCTGCTGTAATGGCCACGAGGACCGCTGTTCTACCTGGCGTC 960
Db      187510 TCGATTATGCGCGCTGCTGTAATGGCCACGAGGACCGCTGTTCTACCTGGCGTC 187451

Qy      961  ACCGGCGCAACAAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGACCGTGGAC 1020
Db      187450 ACCGGCGCAACAAAGAAAGCGGCTCTATGAAGATGGCGTGGCAAGACCGTGGAC 187391

Qy      1021  AACACGATATGATGACCAAGCTGGAAGACGATTCGTGCGAAGCCAGTCAGTGGAC 1080
Db      187390 AACACGATATGATGACCAAGCTGGAAGACGATTCGTGCGAAGCCAGTCAGTGGAC 187331

Qy      1081  GAAGCGCGTGAATTGACGCTTACAGGTTGAATAATA 1119
Db      187330 GAAGCGCGTGAATTGACGCTTACAGGTTGAATAATA 187292

RESULT 10
LOCUS      AY033515 1119 bp DNA linear BCT 26-DEC-2001
DEFINITION Escherichia coli 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
            synthase (ispG) gene, complete cds.
ACCESSION  AY033515
VERSION     AY033515.1 GI:17978527
KEYWORDS   Escherichia coli
SOURCE      Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1 (bases 1 to 1119)
            Hecht,S., Eisenreich,W., Adam,P., Amslinger,S., Kis,K., Bachner,A.,
            Arigoni,D. and Roidich,F.
            Studies on the nonmevalonate pathway to terpenes: The role of the
            Gcpe (ispG) protein
            JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14837-14842 (2001)
            PUBMED   11752431
AUTHORS     2 (bases 1 to 1119)
            Roidich,F., Hecht,S., Adam,P., Bachner,A. and Eisenreich,W.
            Direct Submision
            Submitted (27-Apr-2001) Institut fuer Organische Chemie und
            Biochemie, Lehrstuhl III, Technische Universitaet Muenchen,
            Lichtenberg Strasse 4, Garching D-85747, Germany
FEATURES
            source
            1..1119
            /organism="Escherichia coli"
            /mol_type="genomic DNA"
            /db_xref="taxon:562"
            1..1119
            /gene="ispG"
            1..1119
            /gene="ispG"
            /function="converts 2C-methyl-D-erythritol
            2,4-cyclodiphosphate into 1-hydroxy-2-methyl-2-(E)-butenyl
            4-diphosphate, last but one step of the
            mevalonate-independent pathway for isoprenoid
            biosynthesis"
            /note="ispG"
            /codon_start=1
            /transl_table=1
```

gene	complement (163. .1383) /gene="fabB" /locus_tag="c2869" complement (163. .1383) /gene="fabB" /locus_tag="c2869" /EC_number="2.3.1.41" /function="enzyme; Fatty acid and phosphatidic acid biosynthesis" /note="Escherichia coli K-12 ortholog: b2323; Escherichia coli O157:H7 ortholog: z3586"
CDS	/transl_table=1 /product="3-oxoacyl-[acyl-carrier-protein] synthase I" /protein_id="AA081319.1" /db_xref="GI:26109117" /translation="MKSAVITGLGIVSSINNOQEVLASLRBSRGITTSQELKDSGM RSHVGNVVKLDITGLIDRKVRFMSDASIFAFLSMEQIADAGLSPEAYONPRVGLI AGSGGSPRFQVFGADAMRGRGLKAVGPVYVKAMSGVSACTATPFKIHGVYIS SACATSAHICIGNAVEIOIGKODIVFAGGEBELCEWACEFDMALSTKYNDPEKA SRVYDARDGSPVLAGGSGVYVELHIALRGAIYAEIVYGATSGDAMVPSGEG AYRQMKAMHGVDPIDYLNSHGTSITVGVKELAIRVEFGDKSPATISATKMTAGS LGAGVGEAIYSLMLEHGFVADSINIELEDEQAAGLIVTETTDRELTTVMNSFGF GGTNATLMRKLD"
gene	1482. .3548 /gene="yfcK" /locus_tag="c2870" 1482. .3548 /gene="yfcK" /locus_tag="c2870" /function="putative" /note="Escherichia coli K-12 ortholog: b2324; Escherichia coli O157:H7 ortholog: z3587"
CDS	/transl_table=1 /codon_start=1 /product="Hypothetical protein yfcK" /protein_id="AA081320.1" /db_xref="GI:26109118" /translation="MRSITHTAKIEPPRVRYVYKYSTOPANLENACTPVSDF DDVTFSDNGLEETRYFLGSHLEAFPPHPHPLFVASGFGTGNFTTLWQAFQ FRAHPQAOLOQLHFIISFEKPEPLTRADLAHOMPELPAWAOLOQWPLPGCHR LLDEGHVTLDLWFGDINELTSQLDSDINOKVDAMPJGAPAKNPDMTQNLFNMA RLARPGSTLATFTSAGVFRGLQEGFTMOKRGKGFREMLCGVMEQTLPPFCSTPM FNMMSGKOEAAITGGGIAASALSLALRRGWOVTLVCADEAPALASGNGROALVPL LSHDEALNRFNSNAFTFARFPDILLPVKFDHMCYTQCGMDEKSHKIAQLMSDL PAILAVAEANAVEQITGVANCSGITVPQGMCLCEMLTRNVELAQOGLQIHOH QLODLSRKDGWMLNFAQDGHASHVVLANGHQISGFSQTSLLPVYSAVGQVSHLPT TPELAKIKOVLICYGYLTPONPANQHHICASVYHSGEETAYSEDDQQORLIDCF PQVOMAKEVSVSGEARGGVCAITRDHLPMVGAVPDYDALTVEYASLAEKKDAVSAF VYDDLPMFALGSRGLCSAPLCAEILAAQSEBPIPMIDASTLAAINENRLWVKLLKG KAYKAG"
gene	complement (3669. .3947) /gene="yfcL" /locus_tag="c2871" complement (3669. .3947) /gene="yfcL" /locus_tag="c2871" /note="Escherichia coli K-12 ortholog: b2325; Escherichia coli O157:H7 ortholog: z3588"
CDS	/transl_table=1 /codon_start=1 /product="Hypothetical protein yfcL" /protein_id="AA081321.1" /db_xref="GI:26109119" /translation="MIAFESRIIALIDGVNDAHSDDELFSAGYLGRHLTLTAIELES GDHSAQAVHTVVSQLEKAIAGAEISLPDOALVTDMWEHLFOASQO"
gene	complement (3981. .4529) /gene="yfcM" /locus_tag="c2872" complement (3981. .4529) /gene="yfcM" /locus_tag="c2872" /function="putative"
CDS	
gene	/note="Escherichia coli K-12 ortholog: b2326; Escherichia coli O157:H7 ortholog: z3589"
CDS	/codon_start=1 /transl_table=1 /product="Hypothetical protein yfcM" /protein_id="AA081322.1" /db_xref="GI:26109120" /translation="MNSTHVEQLIEIFNSCFDEFNTRLIKGDEPIYLPADAEPY NRIVFAHGVSAAIHSHWCIGAKARBLVDFGWYCPDGDADQOSQFQEVKAPQ AIDMLFCVAGAPFVUSCONLEGDPEPRVYQRVHAQVMYLIANGIDEPARAFIKA LQNYHTPELTLNEQFWEPRALN"
gene	complement (4529. .5338) /gene="yfcA" /locus_tag="c2873" complement (4529. .5338) /gene="yfcA" /locus_tag="c2873" /function="putative structure" /note="Escherichia coli K-12 ortholog: b2327; Escherichia coli O157:H7 ortholog: z3590"
CDS	/codon_start=1 /transl_table=1 /product="Hypothetical protein yfcA" /protein_id="AA081323.1" /db_xref="GI:26109121" /translation="METSLSFMSVPLLGVLFPVAMLAGFIDSIAGGGGLTTPALM AAGSPANLANTNKLQACGGSATVTFRRRVVSLSDCKINIMTFVMSGGALLVO YVQADYLRQILPIVLCIGLFLPLMPKLGSEORORMYSLPRALLAGGCVGYDFG PAAGSFYALAFVTLGCFNLAKATAHAKLNATSNIGGLFTLDGKVLVATGNTVWG QELGARMGSRVLVSKQKLRPMIVYVSAMSKLLYDSHGQETLHMLQMN"
gene	complement (5338. .6162) /gene="mepA" /locus_tag="c2874" complement (5338. .6162) /gene="mepA" /locus_tag="c2874" /function="enzyme; Murein sacculus, peptidoglycan"
CDS	/transl_table=1 /codon_start=1 /product="Penicillin-insensitive murein endopeptidase precursor" /protein_id="AA081324.1" /db_xref="GI:26109122" /translation="MNKTIALLALIASASIAATFWQKITQPVGSAQSGSPNSGC IYGADLPLOSHEVQVMDRRYFQHPDLVNFIDRLSRQVSNLGMGTVLIGMGMPA GGRFNGHSHQGLDVIDFLQLPKTRWTSQDLRLPQALDLYSRDGHVPAALMKPEI FSLIKLAQDKDVTRFVNPAIKQOLCLDAGDRDRLKRVRFWFOHRAHMYLRCPSA DLECEQDPLPPRGCGGAELOSWEFPRKRGTTKPEKTRPPPLPSCQALLDEHVI"
gene	complement (6166. .7254) /gene="aroC" /locus_tag="c2875" complement (6166. .7254) /gene="aroC" /locus_tag="c2875" /EC_number="4.2.3.5" /function="enzyme; Amino acid biosynthesis: Chorismate"
CDS	/note="Escherichia coli K-12 ortholog: b2329; Escherichia coli O157:H7 ortholog: z3592"
gene	/transl_table=1 /product="Chorismate synthase" /protein_id="AA081325.1" /db_xref="GI:26109123" /translation="MMAAGNTIGQLFRVTFEGSHGLACIIVGVPPGPIPLTEADLOH DDBRRPGTSRTTORBPDQVILISGVEGYTGSILLENTDORSODYSALIDV FRPGHADYVBOKVGIRDYRGGRSARBARVAAAGLAKKYLAKEIETRGCLTO MEDIPLEIKDSQVEQNPFCPPDKIDLDELMLRLKKEGSGIAKYTVVAGSVAG LGEVPRDLADIDIAHLKMSINAVKGEIDGDFVALROSQRNDEITKGFOSNHAG IIGGSSGOQIIAHMALKPTSSITVPGRTINFGEEVEMITGQRHDPVCVIRAVPIAE AMALIVMLDHLFROAQNADVKTDIPRW"

repeat_region	10137..10170	ICOVKGEKGI\$FTTHIAFPYERRMGGFLRDNFKQNRNII"
gene	/note="REP (repetitive extragenic palindromic) element, contains 1 REP sequence"	
gene	/complement(10179..11462)	
gene	/gene="pepB"	
gene	/locus_tag="S2742"	
gene	/complement(10179..11462)	
gene	/gene="pepB"	
gene	/locus_tag="S2742"	
gene	/function="putative enzyme; Not classified"	
gene	/note="residues 1 to 427 of 427 are 92.50 pct identical to residues 30 to 456 of 456 from Escherichia coli K-12 : B2523"	
gene	/codon_start=1	
gene	/transl_table=1	
gene	/product="putative peptidase"	
gene	/protein_id="AAP17894.1"	
gene	/db_xref="GI:30042169"	
gene	/translation="WTEAMKTTLSQPADNAWGEKATYSINDGITLHNGADLGLILQRALIDGLIGIKHVQLSGBGMDADRCWAFWQGIKGPGRKVRYPDIDDAQRELEINRLMIDWDRDITINAPAEELGSQLQARVADI\$SNVAGDRVYRRTTKEDLAEQYMGLHGTGSEERSPVLLALADNPYDGKEAPYALCLVKGITFDSGGYSIKQIAPMDSKEDMTVAATGVALAPAIMRGLNKRKVLFLCCANLIL\$GNAFKGLDIITRYNGKIVVMNTYDABGRTVADGLIDASAKKPEMIIIDATLNGAATGALGVNHALFSPDDLAGLLASASQENBEPFRLPLAEFHRSQLPSNPFELNNTGSAAPYPGASTAAGFLSHFVENVQCKLHIDISATYRRAPAEQW\$AGATGLGVRTANLTLTA"	
gene	/complement(11521..11721)	
gene	/gene="yfhj"	
gene	/locus_tag="S2743"	
Query Match	98.2%; Score 1098.4; DB 1; Length 290380;	
Matches 1105; Conservative 0; Mismatches 11; Indels 0; Gaps 0;		
1	ATGCATACCGAGCGCCCAATTCACAGTGAAGAAATCAACAGGATTTACGTTGGGAATGTG	60
1209	ATGCATACCGAGCGCCCAATTCACAGTGAAGAAATCAACAGGATTTACGTTGGGAATGTG	1150
61	CGGATTTGGCGATGTGTCTCCATCGCCGTACAGTCCATGACCAATACGCGTAGCAGACAG	120
1149	CGGATTTGGCGATGTGTCTCCATCGCCGTACAGTCCATGACCAATACGCGTAGCAGACAG	1090
121	GTGCAAGCAACGCGTCAATCAATCAAGCGCGTGAACGCGTTGGCGGTGATATCTGCGGT	180
1089	GTGCAAGCAACGCGTCAATCAATCAAGCGCGTGAACGCGTTGGCGGTGATATCTGCGGT	1030
181	GTATCCGATACCGAGCGATGAGCGCGGACAGCGTTCAACCTCATCAACACAGAGTTTAA	240
1029	GTCTCCGATACCGAGCGATGAGCGCGGACAGCGTTCAACCTCATCAACACAGAGTTTAA	970
241	GTGCGCGTGTGGCGTCAATCAATCAATCGATTCGATTCGATTCGATTCGATTCGATTCG	300
969	GTGCGCGTGTGGCGTCAATCAATCAATCGATTCGATTCGATTCGATTCGATTCGATTCG	910
301	GGCGTCAATGTCTGCGTATTAACTTGGCAATATCGGTAATGGAAGACGATTCGATG	360
909	GGCGTCAATGTCTGCGTATTAACTTGGCAATATCGGTAATGGAAGACGATTCGATG	850
361	GTGCTTGAATGTGGCGCGGATTAATCAATTCGATTCGATTCGATTCGATTCGATTCG	420
849	GTGCTTGAATGTGGCGCGGATTAATCAATTCGATTCGATTCGATTCGATTCGATTCG	790
421	CTGGAAAAAGATCTGCAAGAAAGATATGGCAACCGACGCGGCGTTGCTGGAATCT	480
789	CTGGAAAAAGATCTGCAAGAAAGATATGGCAACCGACGCGGCGTTGCTGGAATCT	730
481	GCATATGCTCATGTTCATCATCTGATGCGCTGAACTTCGATTCGATTCGATTCGATTCG	540
729	GCATATGCTCATGTTCATCATCTGATGCGCTGAACTTCGATTCGATTCGATTCGATTCG	670
541	AAAGGCTTGAACGCTTCTCGCTGTTGAGTCTTATCGTTTGTGCGCAAAACAGATCGAT	600

Db	669	AAAGGCTGACGCTTCTCTCGCTGTTGAGTCTTATCGTTTGCTGCGAAAACAGTCAT	610
Qy	601	CAGCGCTTGATCTTGGGGATCACCGAGCCGGTGTGCGGACGCGGGACAGTAAATTC	660
Db	609	CAGCGCTGATCTGGGGATCACCGAGCCGGTGTGCGGACGCGGGACAGTAAATTC	550
Qy	661	GCAATTGGTTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	720
Db	549	GCCATTGGTTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	490
Qy	721	GCGGCGCATCCGCTGCAAGAGATCAAAAGTCGGTTTCGATTATTTTGAATTCGCTG	780
Db	489	GCGGCGCATCCGCTGCAAGAGATCAAAAGTCGGTTTCGATTATTTTGAATTCGCTG	430
Qy	781	CGTTGCGAGGAGATCAACTTCATGCTGCTGCCCGACCTGTTGCGCTCAGGAATTGAT	840
Db	429	CGTTGCGAGGAGATCAACTTCATGCTGCTGCCCGACCTGTTGCGCTCAGGAATTGAT	370
Qy	841	ATCGGTACGGTTACCGCGCTGAGACACCGCTGAAAGATATCATCTCCATGGACGTT	900
Db	369	ATCGGTACGGTTACCGCGCTGAGACACCGCTGAAAGATATCATCTCCATGGACGTT	310
Qy	901	TCGATTATCGCTCGTGTGTAATGAGCCAGGTGAGCGCTGCTTCTACACTCGCGCTC	960
Db	309	TCGATTATCGCTCGTGTGTAATGAGCCAGGTGAGCGCTGCTTCTACACTCGCGCTC	250
Qy	961	ACGCGCGGCAACAAGAAAAGCGGCTCTATGAGATGGCGTGGCAAGACGCTGGAC	1020
Db	249	ACGCGCGGCAACAAGAAAAGCGGCTCTATGAGATGGCGTGGCAAGACGCTGGAC	190
Qy	1021	AACAAACGATATGATCGACCACTGTGGAAGACAGCATTCGTGGAAAGCCAGTCAGT	1080
Db	189	AACAAACGATATGATCGACCACTGTGGAAGACAGCATTCGTGGAAAGCCAGTCAGT	130
Qy	1081	GAGCGCGCTCGAATTGACGTTCCAGCAGGTTGAAAA	116
Db	129	GAGCGCGCTCGAATTGACGTTCCAGCAGGTTGAAAA	94
RESULT 9			
AE016764/c			
LOCUS	AE016764	300099 bp	DNA linear BCT 24-DEC-2002
DEFINITION	Escherichia coli CFT073 section 10 of 18 of the complete genome.		
ACCESSION	AE016764	AE014075	
VERSION	AE016764.1	GI:26109116	
KEYWORDS			
SOURCE			
ORGANISM	Escherichia coli CFT073		
	Escherichia coli CFT073		
	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
	Enterobacteriaceae; Escherichia.		
REFERENCE	1 (bases 1 to 300099)		
AUTHORS	Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,		
	Reiko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,		
	Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,		
	Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.		
	Extensive Mosaic Structure Revealed by the Complete Genome Sequence		
	of Uropathogenic Escherichia coli		
	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)		
TITLE			
JOURNAL	PUBMED		
REFERENCE	12471157		
AUTHORS	2 (bases 1 to 300099)		
	Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,		
	Raeko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,		
	Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,		
	Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-JUN-2002) Genetics Laboratory, University of		
FEATURES	Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA		
source	Location/Qualifiers		
	1..300099		
	/organism="Escherichia coli CFT073"		
	/mol_type="genomic DNA"		
	/strain="CFT073"		
	/db_xref="taxon:199310"		

TITLE Direct Submission
JOURNAL Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
Source
1. .290380
/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"
/serotype="2a"
/db_xref="taxon:198215"
complement (82. .1209)
/gene="gcpE"
/locus_tag="S2733"
complement (82. .1209)
/gene="gcpE"
/locus_tag="S2733"
/note="residues 1 to 372 of 375 are 96.23 pct identical to residues 1 to 372 from Escherichia coli K-12 : B2515"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAPI7888.1"
/db_xref="GI:30042163"
/translation="MHNQAPIORRKSRIYVGNVPIGDGAPIAVOSMTNRTTDEAT VNOIKLERYGADIVRVSPMDAAEFKLIKORVNVPLVDIHPDRIALKVAEYGV DCRIRNGINIGNERIRRMVDCARDKNIPIRIGNAGSLKDKOEKXGPTPALLES AMHVDHLDRNFQFKVKSADSVPLAVESYRLAKQIDQPLHLGITTEGARSQAV KSAIGLGLLESGIDTLRVSLADPVEELKVGEDILKSLIRSRGINFLACPTCSNQ EFPVIGTVNLEGLELIEDITPMQVSLIGCVNNGEALVSTLGVGNGKSGLYEDGV RKRLDNDMDIDLEARIKASOLSDBARRIDVQGVKITT"
complement (1236. .2249)
/gene="yfgA"
/locus_tag="S2734"
complement (1236. .2249)
/gene="yfgA"
/locus_tag="S2734"
/function="putative membrane; Not classified"
/note="residues 1 to 337 of 337 are 80.71 pct identical to residues 1 to 337 of 337 from Escherichia coli K-12 : B2516"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="AAPI7889.1"
/db_xref="GI:30042164"
/translation="MNTTEATHDQNEALTGTGARLNAREQLGSLQAAVERLCIKSVT RDIEDKAPADLASTFLRGYIRSYARLVHIDPEELPGLKQAPLRAKVAPQSFSL GRRKKRDGWLMTFTLVLFVIGLSGAMMMQDHKAQOEITTMADQSSAELSSNSBO GGVPLNTSTTTPATSTSPASVDTTATNTQTPAVTAPAPAVDPOONAVVSPQNAV DTATPTVPTAATTPDGAAPLPTDQAGVTTTAADPNALVMNFTADCHLEVTDAQKELF SGWQRKQGNLNTGQAPYKLTIGAPAAVQIQGKPVLSRFTRTNOVARTLNAEOS PAQ"
complement (2534. .3688)
/gene="yfgB"
/locus_tag="S2735"
complement (2534. .3688)
/gene="yfgB"
/locus_tag="S2735"
/note="residues 1 to 384 of 384 are 99.47 pct identical to residues 1 to 384 of 384 from Escherichia coli K-12 : B2517"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAPI7890.1"
/db_xref="GI:30042165"
/translation="MSQIVTPENVTTRKDGKINLIDNRQMEFEKDLGKTPRADO VMKMYHYCCDNPEMTDINKVLAGKLEKAEVETAPRVVEQSSDDGTITMAIANGQ RMTVYIPEDRATLVCSSOVGALBECKFCSTAOQGNRNDRVSEITGVWRKAKTGV AAKVTGQRPITNNVVMGMEGPLNLNNVPMAMETMDDFGSLSKRVTLSTGSRVPA LDKIGMDIVALAISLAHNPDEIDELVPIINKKNTITFLAAVARYSEKSNANGCVT IEYVMDHVDNGTEHAHQALAEILKDTPTCKINLIIPWNPFPDAPVGRSSNSRIIDFSKVL

MSYGFITVRKTRGDDIDACQGLAGVIDRTRKTLRKMOGEAIDIKAV"
complement (3838. .4269)
/gene="ndk"
/locus_tag="S2736"
complement (3838. .4269)
/gene="ndk"
/locus_tag="S2736"
/function="enzyme; Purine ribonucleotide biosynthesis"
/note="residues 1 to 143 of 143 are 100.00 pct identical to residues 1 to 143 of 143 from Escherichia coli K-12 : B2518"
/codon_start=1
/transl_table=1
/product="nucleoside diphosphate kinase"
/protein_id="AAPI7891.1"
/db_xref="GI:30042166"
/translation="MAIERFSTIKENAVAKVIGNIFARFEAGKIVGTQMLHLTV EQARKFYAHDKRPFPGDLVFERMTSGPIVSVLEGNAVQRHRDLGATNPANALAGT LRADYADSLTENGTHGSDSVESAAREIAVFPGEVCPRTT"
complement (4418. .6517)
/gene="pbpC"
/locus_tag="S2737"
/note="truncated"
/pseudo
6514. .7961
/note="insertion element"
/insertion_seq="ISSf14"
complement (6534. .7730)
/locus_tag="S2738"
complement (6534. .7730)
/locus_tag="S2738"
/function="IS, phage, Tn; Transposon-related functions"
/note="residues 1 to 398 of 398 are 90.95 pct identical to residues 1 to 398 of 398 from GenPept : >db|AKI8438.1|AF348706.127 (AF348706) putative transposase for IS110 [Shigella flexneri]"
/codon_start=1
/transl_table=1
/product="ISSf14 orf"
/protein_id="AAPI7892.1"
/db_xref="GI:30042167"
/translation="MTSSDYBSVQVFIGVDVGDTHAAVAINRSGRLFDKALPNDE NKRLSLIDLKQHQIILVDDQPAITGALPVAVRSEGVYGLPGLAMRILADLHAG EAKTVRDAAITIABARTLPHALRTLKLEDOIAELSMCGFPDDLAAGTOSNRIR GULTQHPRPBVLGRLRHPAVLDLQVYSPSEKIASIGKKLAAQLCKLARLGR LADIDQALAEQTVVGGTNAAVVPRALQITIRKQDEVALEVEGRVLAHPIYP VLTSMGVAVRTAARLLTEVACRAFAAHLAAVAGLAVTRSGSSIGEHPSRGN KALKRALFISAPAAALRDPUSRAYYTRKMSQGRHQAALALARRCDVLFAMWRDGT FYPOGS"
7960. .8442
/gene="aseA"
/locus_tag="S2740"
/note="truncated"
/pseudo
9261. .10037
/gene="aseB"
/locus_tag="S2741"
complement (9261. .10037)
/gene="aseB"
/locus_tag="S2741"
/function="phenotype; Not classified"
/note="residues 1 to 258 of 258 are 94.96 pct identical to residues 4 to 261 of 261 from Escherichia coli K-12 : B2522"
/codon_start=1
/transl_table=1
/product="enhancer of serine sensitivity"
/protein_id="AAPI7893.1"
/db_xref="GI:30042168"
/translation="MSETKLEBDLLEKATERRHAPARFRTLLSETWVPCRAAGE AVEDSALDQHEKEDGTSVLPFTSLDLOQAVDEQAIVMPVPRTLFEMTLGSL FLNAKLPTKEKMPREISLLIGEGNPLSSOELIEGSELISVVAEPPAOMDSLIT LFKTIPVRAFAFICSIKENEAQPNILLIGIEADGDIIEIIGATGVSATDTLPDGEID

CDS

locus_tag="SF2561"
 complement(6515..7642)
 /gene="gcpB"
 /locus_tag="SF2561"
 /note="Residues 1 to 372 of 375 are 99 pct identical to residues 1 to 372 of a 372 aa protein from *Escherichia coli* O157:H7 EDL933 ref: NP_289068.1"
 /codon_start=1
 /transl_table=11
 /product="orf, conserved hypothetical protein"
 /protein_id="AAM4061.1"
 /db_xref="GI:24052935"
 /translation="MHNQAPIORRKRSTRIVGNVPIGDGPAIVASMTNTRTVDVENTVNOIKALERVGADIVRSVPTMDAAEAFKIKORVNPVADIHFDYRIALKYAEYGDCKRINQNGIERIRIMVDCARDKNIPIRIGNAGSLEKDIQEKGEPTPOLAESAMHVDLIDNLSEIDQFVSVKASDVPLAVSRYLAKOIPHLGITTEAGSGASVKSALIGLILSEIGDTLRVSLAADPREEIKVGFIDIKSLIRISRGINFACPTCSROEPVIGTVNALQORLEDIITPMDSVLIIGVNVGPEALVSTIGTVTGKSKSLYEDGV RKDRLNDMDIMDLQEARIRAKASQLEARRIDVQVQKILIT"
 /gene="yfgA"
 complement(7669..8682)
 /locus_tag="SF2562"
 complement(7669..8682)
 /gene="yfgA"
 /locus_tag="SF2562"
 /note="Residues 1 to 337 of 337 are 99 pct identical to residues 1 to 337 of a 337 aa protein from *Escherichia coli* O157:H7 ref: NP_311405.1"
 /codon_start=1
 /transl_table=11
 /product="putative membrane protein"
 /protein_id="AAM4062.1"
 /db_xref="GI:24052936"
 /translation="NMTREATHPDONEALITTGARLRNAREQIGLSQAVARLCLKVSTV RDIEDCAPADLSTPLRGITRSYARVHLPBELLPGELKQAPLRAPKAPWQSTSL GSKRRKDGMLFTWLVLFFVNLGSGAMWQDKAQOEIITMADQSSAESLSNEO GQSVPLNTSTTTPATSTPASPVDITATQTPAVTAAPAPADPOONAAVSPSOANV DTAATPVPATTPDGAAPLETPDQAVTTPADPNALVMVFADCMLEVTDAFGKLLF

Query Match 98.2% Score 1098.4; DB 1; Length 10225;
 Best Local Similarity 99.0% Pred. No. 1,1e-260;
 Matches 1105; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 1 ATGCATACACAGGCTCCATTCAACGAGAAATCAACAGCATTTTACGTTGGGAATGTG 60
 DB 7642 ATGCATACACAGGCTCCATTCAACGAGAAATCAACAGCATTTTACGTTGGGAATGTG 7583
 QY 61 CCGATTGGCGATGGTGTCTCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
 DB 7582 CCGATTGGCGATGGTGTCTCCATCGCCGTACAGTCCATGACCAATACGCGTACGACAGAT 7523
 QY 121 GTGGAAGCAACGGTCAATCAATCAAGCGCGCTGGAACGCGTGGCGTGAATATCTCCGT 180
 DB 7522 GTGGAAGCAACGGTCAATCAATCAAGCGCGCTGGAACGCGTGGCGTGAATATCTCCGT 7463
 QY 181 GTATCGGTACGACGATGACGGGGGAGAGCGTTCAACTCAATCAAGAGCGTTAAC 240
 DB 7462 GTATCGGTACGACGATGACGGGGGAGAGCGTTCAACTCAATCAAGAGCGTTAAC 7403
 QY 241 GTGCCGCTGGTGGCTGACATCACTTCGACTATCGATTGCGGTGGAAGTAGGGAATAC 300
 DB 7402 GTGCCGCTGGTGGCTGACATCACTTCGACTATCGATTGCGGTGGAAGTAGGGAATAC 7343
 QY 301 GGGCTGATTTGTCTGGTATTAACCTCGCAATATCGGTATGAAGAGCTATTCCGATG 360
 DB 7342 GGGCTGATTTGTCTGGTATTAACCTCGGTATTAATCGGTATGAAGAGCTATTCCGATG 7283
 QY 361 GTGGTGAATGTGTCGGCGCGATTAACATTCGATCCGATTCGTTAGCGTTAACGCCGATCG 420
 DB 7282 GTGGTGAATGTGTCGGCGCGATTAACATTCGATTCGTTAGCGTTAACGCCGATCG 7223
 QY 421 CTGGAAGAAAGATCTGCAAGAAAGATGCGAAGCCGACCGGACGCGTGTGTAATCT 480

DB 7222 CTGGAAGAAAGATCTGCAAGAAAGATGCGAAGCCGACCGCGCTGTGTAATCC 7163
 QY 481 GCCATGCGTATGTTGATCATTCGATCGCGCTGAACCTTCGATCAAGTACAGCGTG 540
 DB 7162 GCCATGCGTATGTTGATCATTCGATCGCGCTGAACCTTCGATCAAGTACAGCGTG 7103
 QY 541 AAGCGTCTGACGCTCTTCTCGTCTGTGTAGTCTTATCGTTTCTGCAAAACAGATCAT 600
 DB 7102 AAGCGTCTGACGCTCTTCTCGTCTGTGTAGTCTTATCGTTTCTGCAAAACAGATCAT 7043
 QY 601 CAGCGTTCATTTGGGGATCACGAAAGCGGTGTGGCGGACGCGGGCAGTAAATCC 660
 DB 7042 CAGCGTTCATTTGGGGATCACGAAAGCGGTGTGGCGGACGCGGGCAGTAAATCC 6983
 QY 661 GCCATGTTTGGTGTGCTGCTGTGTAAGCATGGGGAACGCTGGCGGATGCGTG 720
 DB 6982 GCCATGTTTGGTGTGCTGCTGTGTAAGCATGGGGAACGCTGGCGGATGCGTG 6923
 QY 721 GCGGCGATCCGCTCGAAGAGATCAAGTCCGTTTGGATATTTGAAATCGCTGCTATC 780
 DB 6922 GCGGCGATCCGCTCGAAGAGATCAAGTCCGTTTGGATATTTGAAATCGCTGCTATC 6863
 QY 781 CGTTCGAGAGGATCAACTTCATCGCTCGCGACCTGTTCCGCTGAGAAATTTGATTT 840
 DB 6862 CGTTCGAGAGGATCAACTTCATCGCTCGCGACCTGTTCCGCTGAGAAATTTGATTT 6803
 QY 841 ATCGGATCGTTTAAAGCGCTGAGCAAGCGCTGGAAGATATCATCTCGATGAGCGTT 900
 DB 6802 ATCGGATCGTTTAAAGCGCTGAGCAAGCGCTGGAAGATATCATCTCGATGAGCGTT 6743
 QY 901 TCGATTATCGGCTGCTGTGTAATGCGCCAGGTGAGCGCGTGTCTTCACTCGGCGTC 960
 DB 6742 TCGATTATCGGCTGCTGTGTAATGCGCCAGGTGAGCGCGTGTCTTCACTCGGCGTC 6683
 QY 961 ACCGCGGCAACAAGAAAGCGGCTCTTATGAAGATGCGGTGCGCAAGACGCTTGAGAC 1020
 DB 6682 ACCGCGGCAACAAGAAAGCGGCTCTTATGAAGATGCGGTGCGCAAGACGCTTGAGAC 6623
 QY 1021 AACAAAGATATGATGACAGCGTGAAGAGCAAGCATTCGTCGGAACAGTCAAGTCAAG 1080
 DB 6622 AACAAAGATATGATGACAGCGTGAAGAGCAAGCATTCGTCGGAACAGTCAAGTCAAG 6563
 QY 1081 GAAGCGCTGCAATGACGTTTCAGCAGTTGAAAAA 1116
 DB 6562 GAAGCGCTGCAATGACGTTTCAGCAGTTGAAAAA 6527

RESULT 8
 AE016987/c
 LOCUS
 Shigella flexneri 2a str. 2457T section 10 of 16 of the complete genome.
 ACCESSION
 AE016987
 VERSION
 AE016987.1 GI:30042162
 KEYWORDS
 SOURCE
 ORGANISM
 Shigella flexneri 2a str. 2457T
 Shigella flexneri 2a str. 2457T
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
 REFERENCE
 1 (bases 1 to 290380)
 Fournier, G., Mayhew, G.F., Burland, V., Venkatesan, M.M., Deng, W., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.
 Complete Genome Sequence and Comparative Genomics of *Shigella flexneri* Serotype 2a Strain 2457T
 Infect. Immun. 71 (5), 2775-2786 (2003)
 JOURNAL
 PUBMED
 12704152
 2 (bases 1 to 290380)
 Fournier, G., Mayhew, G.F., Burland, V., Venkatesan, M.M., Deng, W., Wei, J., Goldberg, M.B., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C. and Blattner, F.R.

TITLE Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity through comparison with genomes of *Escherichia coli* K12 and O157

JOURNAL Nucleic Acids Res. 30 (20), 4432-4441 (2002)

PUBMED 12384590

REFERENCE (bases 1 to 10225)

AUTHORS Jin, Q., Shen, Y., Wang, J.H., Liu, H., Yang, J., Yang, F., Zhang, X.B., Zhang, J.Y., Yang, G.W., Wu, H.T., Dong, J., Sun, L.L., Xue, Y., Zhao, A.L., Gao, Y.S., Zhu, J.P., Chen, S.X., Yao, Z.J., Wang, Y., Lu, W.C., Qiang, B.Q., Wen, Y.M. and Hou, Y.D.

TITLE Direct Submission

JOURNAL Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P.R. China

FEATURES

source Location/Qualifiers

1. 10225

/organism="Shigella flexneri 2a str. 301"

/mol_type="genomic DNA"

/strain="301"

/serotype="2a"

/db_xref="taxon:198214"

148. 1440

/gene="xsea"

/locus_tag="SF2555"

148. 1440

/gene="xsea"

/locus_tag="SF2555"

/note="Residues 1 to 430 of 430 are 97 pct identical to residues 27 to 456 of a 456 aa protein from *Escherichia coli* K12 ref: NP_417004.1"

/codon_start=1

/transl_table=1

/product="exonuclease VII, large subunit"

/protein_id="AA04055.1"

/db_xref="GI:24052929"

/translation="MGQVVISGEISNFTOPASGHWFTLKDDTAQVRCAMFRNSNRVTPPHOGQOVLVRANITLIEPRDGYOIVESMOPAGGLQOKEOLKALQECLEDOCKKPLSPSPACVGVTSKTAALHILVLRKRPDLPIIIVPTSVGGDAPGQIVRAELANORNECVLIVRGSGSLEDIMSFNDRVRAIPASRIPIVSAAGHETDVIADFDVADRAPTPSAAEVSVRNOQELIRVOQSAORLEMDYLLARTIRFPQIHHRLOOQHPLRLAQQTMLERLOKMSFALENQLKRAQGOQRLQRLNQMPQKIHRTOTRIQOLEVRLAEILRAQLSATRRFENAVTHLEAVSPSLTARGYSTTATADGNVLYKKVQKVGEMLTTRLEDGWTIESEVKNIQPVKSKKKVH"

complement(1437..1688)

/gene="yfgu"

/locus_tag="SF2556"

complement(1437..1688)

/gene="yfgu"

/locus_tag="SF2556"

/note="Residues 1 to 83 of 83 are 97 pct identical to residues 1 to 83 of a 83 aa protein from *Escherichia coli* K12 ref: NP_417005.1"

/codon_start=1

/transl_table=1

/product="orf, conserved hypothetical protein"

/protein_id="AA04056.1"

/db_xref="GI:24052930"

/translation="MNVGMAVGIGIHMEIHCPCQHYLDQNGHRCPSGGEIEMKALCPDCHQPLQVLAKGAVDYFCOHGLISKKEFVLA"

complement(1721..3332)

/gene="yfgk"

/locus_tag="SF2557"

complement(1721..3332)

/gene="yfgk"

/locus_tag="SF2557"

/note="Residues 1 to 503 of 503 are 99 pct identical to residues 1 to 503 of a 503 aa protein from *Escherichia coli* O157:H7 EDL933 ref: NP_289064.1"

/codon_start=1

/transl_table=1

/product="putative GTP-binding factor"

/protein_id="AA04057.1"

/db_xref="GI:24052931"

/translation="NRCLMITYKNEALNMPVVALYGRPNVKGSTLFFNRLTRDALVADPPGLRDRKVRARIEGEFFICIDTGGDGEDVETMAEQSLAIIEADVLMVDAKGLMPADDEAIKALIRSRKEPTLVYAKTGLDPODAVDFPSGLGEIYVIAASHGKGVSLIETHVILPMMEDIAAPQEVDEDEAYNAQFAEENGEEBDEDDPQSLPIKLATVGRPVKSTLTNRKILGEERVVYDMGTDRSDSYIIMERDGRERYVLITAGVKRGIITDAVEKESVITKLOAIEDANVWLVIDAEGISDQSLGLFPLNSGRSLVIVNKMDSLOEKEQVKETIDRLGIFIDPARVHFIISALGSGVNLFESEVRAVYOSTERVGTSMRLRTIMTVAEDHPPLVGRVRYKLYAHAGVYNPPIVVIHGVQVLDPPSYKRYLMNYPKSLDWNKSGSPIRIOFKEGENYANKRTVLTPTQMRKKRLMKHLMKK"

complement(3311..4489)

/gene="yfgl"

/locus_tag="SF2558"

complement(3311..4489)

/gene="yfgl"

/locus_tag="SF2558"

/note="Residues 1 to 392 of 392 are 99 pct identical to residues 1 to 392 of a 392 aa protein from *Escherichia coli* K12 ref: NP_417007.1"

/codon_start=1

/transl_table=1

/product="putative dehydrogenase"

/protein_id="AA04058.1"

/db_xref="GI:24052932"

/translation="MQLRLILPGLISVTLISGCSLFNSEEDVYKNSPLPTVENQPTPTAWSTVSGSIGNFYSNHPALADNVVYAAABRAGLVKALNDDKEIWSVLSAEDGWFSKEPALLSGGVTVSGHVYIGSEKAOYVANTSGTAMQTKVAGELSRPVSDGLVLIHNSNQLOALNADADAVKVTYVLMPSLISRESAPIATFGAIVGCGNGRVAVLMEGQOMIMQORISQATGSTIDRLSDVDTTPVYVNGVFPALVANGNTALDRLSGQIMVKRELGVNDFIVDEGNRIYLVDOQDRWALITDGGVILMQSDLLHLLTSPVLYN GMLVDSSEGYLWIMINVEDGRFVAQQKVDSSGFPQEPVADKLLIQAQDGVYSYTR"

complement(4500..5120)

/gene="yfgm"

/locus_tag="SF2559"

complement(4500..5120)

/gene="yfgm"

/locus_tag="SF2559"

/note="Residues 1 to 206 of 206 are 98 pct identical to residues 1 to 206 of a 206 aa protein from *Escherichia coli* K12 ref: NP_417008.1"

/codon_start=1

/transl_table=1

/product="orf, conserved hypothetical protein"

/protein_id="AA04059.1"

/db_xref="GI:24052933"

/translation="MEIYENNDQYEAVERFPFAENGKALAVGILGVALLTGKRYNNSHOVDSARSASLAVYONAVTVASEGKPSIPAAEKFAENKNTYGALSLBLAQOQVQKNELEKAAQLOQGLADTSDENLRKAVINLRAROVOLKQADALKTIDITIKGEGMAIIVADLRGALLSKGDKQCARSAWAGVSDVTPLSEMMQKINMLSI"

complement(5138..6412)

/gene="hiss"

/locus_tag="SF2560"

complement(5138..6412)

/gene="hiss"

/locus_tag="SF2560"

/note="Residues 1 to 424 of 424 are 99 pct identical to residues 1 to 424 of a 424 aa protein from *Escherichia coli* O157:H7 EDL933 ref: NP_289067.1"

/codon_start=1

/transl_table=1

/product="histidine tRNA synthetase"

/protein_id="AA04060.1"

/db_xref="GI:24052934"

/translation="MAKNIQAIRGMANDIYPGETAIMQRIEGLKAVLVSGYSEILPIVEQTPLEFRAIGEIVTVDEKEMVTFEDNKGSLTRPFGTACVAGIEHGLVIOEOLWYIGPWFRRERPOKGRYQFHOQCEVFGDIDAEILMLVRAWRALFIEHVTLEINSIGSLBARAPDALVAFLQHEKLEDEDCKRMYNPLPLVDSKNEVQALINDAPRLGVLDEBSREHPAGLCKLSEAGIAYTVNQRVLRGDIYNRVTFEYVNTSLGSGCTVACAGRYDGLVEOLIGRATPAVAGVAGERIVLLVQAVNPEFKADPVDIVLYASGADTQSAAMALAEIRLDELIGVYKLMTHGGSNFKQAPARADKMARAAVVLGSEBVAAGTAVVQDLRGEQTAVAQDSVAALRLTLTG"

complement(6515..7642)

/gene="gcpf"

REFERENCE
AUTHORS
Bacteri, Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
1 (bases 1 to 13176)
Blattner, F.R., Plunkett, G. III, Bloch, C.A., Berna, N.T., Burland, V., Riley, M., Colado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y.
The complete genome sequence of *Escherichia coli* K-12
Science 277 (5331), 1453-1474 (1997)
97426617
MEDLINE
9278503
2 (bases 1 to 13176)
Blattner, F.R.
Direct Submission
Submitted (16-JUN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
3 (bases 1 to 13176)
Blattner, F.R.
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
4 (bases 1 to 13176)
Plunkett, G. III
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHRR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: mark@amber.gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). ** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES
source
Location/Qualifiers
1. 13176
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
complement(49..81)
/note="factor Sigma70; promoter hns; documented +1 at 263862"

promoter
complement(86..1204)
/gene="gcpB"
/note="synonym: b2515"
complement(86..1204)
/gene="gcpB"
/function="orf; Unknown"

CDS

/note="f372; 100 pct identical to GCPB_ECOLI SW: P27433"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC75568.1"
/db_xref="GI:1788863"
/translation="MNOAPIQRRKSTRIVGNVPIGDGAPIAVQSMINTRTTVEATVNOIKLEKRVGADIVASVPTMDAAEFLLIKQVNVPIVADIHEDYRIALKAEGVDCLRINPQNGNEERIRVMVDCARDNIPRIQVNAQSGLEKLOEKGEPTPALLESARHVDHILNPNFQDKSVKASDVLAESVRLAKOIDOPHIGITTEAGARSGAVKSAIGLILISGIGDTLRLASLADPVEIKGFQDLKSLRISRGINTACTCGRQEPDVGTVALEQRLEDIITPMVDSTIGCVVGPGLAVSTIGVTGNKSGLYLEDGV RKDRLDNMIMDQLEAIRAKASQIDERRIDVQVEK"
complement(1231..2244)
/gene="yfgA"
/note="synonym: b2516"
complement(1231..2244)
/gene="yfgA"
/function="putative membrane, Not classified"
/note="f337; 100 pct identical to YFGA_ECOLI SW: P27434"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAC75569.1"
/db_xref="GI:1788864"
/translation="MNTTEATHDQNEALTTGARLNRNAREQLGSQLVAERLCLKVSTVRIEEDKAPADLASTPFLRGVRSYARLVHPEBELPGEKAPLEAAVAPQOSFLGKRKRKRDGMVMTFTLVLPVYIGSGAMWMDQKAOEITTMAROSSAEISNSQEGSVPLNITTTDPTTSTPPASVDITANTOTPAYTAAPAVDPQQAIVSSQANVDTAATPAPTAATPDGAAPLPDQAGVATPVADPNLVNPFADCVLEVDATGKLP SGWQRDGNLNTGQAPYKLIKIGAPAAVOIYQGKEVDLSRFIRITQVARTLNAOS PAO"
complement(1288..1316)
/gene="yfgA"
/note="factor Sigma70; predicted +1 start at 2639901"
complement(1360..1389)
/gene="yfgA"
/note="factor Sigma70; predicted +1 start at 2639973"
complement(2529..3683)
/gene="yfgB"
/note="synonym: b2517"
complement(2529..3683)
/gene="yfgB"
/function="orf; Unknown"
/note="f384; 100 pct identical to YFGB_ECOLI SW: P36979"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC75570.1"
/db_xref="GI:1788865"
/translation="MSEQLVTPENVTTKDKINLIDLNRQMRFPKDLGKXPFRADQVKMKTHYCDNDEMDTINKYLRLGKLYEAKRABVVEQSSSGCTIKMALAVDO RAEVTVIPEDDRLTLCVSSQVGCALCKFCSTROQGFNNRLRSEIIGVWRKAYGVAATVYQQRIPITVVMWGEPLNLNNVPAMIMDDGFGLSKRVTLSGVPAIDLGLMIVVALAISLAPNDEIRDIVINKKYNIEFLAAVRYLEKSNNAQGVTLIEVVMDDHNDGEHHAOLAEILKCPCKINILPNMPFGAPYGRSSNSRIDFSKVLMSYGFITVIRKTRGGDDIDACGQLADVIDRTKRTLKRMQGAIDIKAV"
complement(3778..3806)
/note="factor Sigma70; predicted +1 start at 2642391"
complement(3786..3814)
/note="factor Sigma70; predicted +1 start at 2642399"
complement(3833..4264)
/gene="ndk"
/note="synonym: b2518"
complement(3833..4264)
/gene="ndk"
/EC_number="2.7.4.6"
/function="enzyme; Purine ribonucleotide biosynthesis"
/note="f143; 100 pct identical to NDK_ECOLI SW: P24233"
/codon_start=1
/transl_table=11
/product="nucleoside diphosphate kinase"
/protein_id="AAC75571.1"

promoter
complement(1231..2244)
/gene="yfgA"
/note="synonym: b2516"
complement(1231..2244)
/gene="yfgA"
/function="putative membrane, Not classified"
/note="f337; 100 pct identical to YFGA_ECOLI SW: P27434"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAC75569.1"
/db_xref="GI:1788864"
/translation="MNTTEATHDQNEALTTGARLNRNAREQLGSQLVAERLCLKVSTVRIEEDKAPADLASTPFLRGVRSYARLVHPEBELPGEKAPLEAAVAPQOSFLGKRKRKRDGMVMTFTLVLPVYIGSGAMWMDQKAOEITTMAROSSAEISNSQEGSVPLNITTTDPTTSTPPASVDITANTOTPAYTAAPAVDPQQAIVSSQANVDTAATPAPTAATPDGAAPLPDQAGVATPVADPNLVNPFADCVLEVDATGKLP SGWQRDGNLNTGQAPYKLIKIGAPAAVOIYQGKEVDLSRFIRITQVARTLNAOS PAO"
complement(1288..1316)
/gene="yfgA"
/note="factor Sigma70; predicted +1 start at 2639901"
complement(1360..1389)
/gene="yfgA"
/note="factor Sigma70; predicted +1 start at 2639973"
complement(2529..3683)
/gene="yfgB"
/note="synonym: b2517"
complement(2529..3683)
/gene="yfgB"
/function="orf; Unknown"
/note="f384; 100 pct identical to YFGB_ECOLI SW: P36979"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC75570.1"
/db_xref="GI:1788865"
/translation="MSEQLVTPENVTTKDKINLIDLNRQMRFPKDLGKXPFRADQVKMKTHYCDNDEMDTINKYLRLGKLYEAKRABVVEQSSSGCTIKMALAVDO RAEVTVIPEDDRLTLCVSSQVGCALCKFCSTROQGFNNRLRSEIIGVWRKAYGVAATVYQQRIPITVVMWGEPLNLNNVPAMIMDDGFGLSKRVTLSGVPAIDLGLMIVVALAISLAPNDEIRDIVINKKYNIEFLAAVRYLEKSNNAQGVTLIEVVMDDHNDGEHHAOLAEILKCPCKINILPNMPFGAPYGRSSNSRIDFSKVLMSYGFITVIRKTRGGDDIDACGQLADVIDRTKRTLKRMQGAIDIKAV"
complement(3778..3806)
/note="factor Sigma70; predicted +1 start at 2642391"
complement(3786..3814)
/note="factor Sigma70; predicted +1 start at 2642399"
complement(3833..4264)
/gene="ndk"
/note="synonym: b2518"
complement(3833..4264)
/gene="ndk"
/EC_number="2.7.4.6"
/function="enzyme; Purine ribonucleotide biosynthesis"
/note="f143; 100 pct identical to NDK_ECOLI SW: P24233"
/codon_start=1
/transl_table=11
/product="nucleoside diphosphate kinase"
/protein_id="AAC75571.1"

promoter
complement(1231..2244)
/gene="yfgA"
/note="synonym: b2516"
complement(1231..2244)
/gene="yfgA"
/function="putative membrane, Not classified"
/note="f337; 100 pct identical to YFGA_ECOLI SW: P27434"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAC75569.1"
/db_xref="GI:1788864"
/translation="MNTTEATHDQNEALTTGARLNRNAREQLGSQLVAERLCLKVSTVRIEEDKAPADLASTPFLRGVRSYARLVHPEBELPGEKAPLEAAVAPQOSFLGKRKRKRDGMVMTFTLVLPVYIGSGAMWMDQKAOEITTMAROSSAEISNSQEGSVPLNITTTDPTTSTPPASVDITANTOTPAYTAAPAVDPQQAIVSSQANVDTAATPAPTAATPDGAAPLPDQAGVATPVADPNLVNPFADCVLEVDATGKLP SGWQRDGNLNTGQAPYKLIKIGAPAAVOIYQGKEVDLSRFIRITQVARTLNAOS PAO"
complement(1288..1316)
/gene="yfgA"
/note="factor Sigma70; predicted +1 start at 2639901"
complement(1360..1389)
/gene="yfgA"
/note="factor Sigma70; predicted +1 start at 2639973"
complement(2529..3683)
/gene="yfgB"
/note="synonym: b2517"
complement(2529..3683)
/gene="yfgB"
/function="orf; Unknown"
/note="f384; 100 pct identical to YFGB_ECOLI SW: P36979"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC75570.1"
/db_xref="GI:1788865"
/translation="MSEQLVTPENVTTKDKINLIDLNRQMRFPKDLGKXPFRADQVKMKTHYCDNDEMDTINKYLRLGKLYEAKRABVVEQSSSGCTIKMALAVDO RAEVTVIPEDDRLTLCVSSQVGCALCKFCSTROQGFNNRLRSEIIGVWRKAYGVAATVYQQRIPITVVMWGEPLNLNNVPAMIMDDGFGLSKRVTLSGVPAIDLGLMIVVALAISLAPNDEIRDIVINKKYNIEFLAAVRYLEKSNNAQGVTLIEVVMDDHNDGEHHAOLAEILKCPCKINILPNMPFGAPYGRSSNSRIDFSKVLMSYGFITVIRKTRGGDDIDACGQLADVIDRTKRTLKRMQGAIDIKAV"
complement(3778..3806)
/note="factor Sigma70; predicted +1 start at 2642391"
complement(3786..3814)
/note="factor Sigma70; predicted +1 start at 2642399"
complement(3833..4264)
/gene="ndk"
/note="synonym: b2518"
complement(3833..4264)
/gene="ndk"
/EC_number="2.7.4.6"
/function="enzyme; Purine ribonucleotide biosynthesis"
/note="f143; 100 pct identical to NDK_ECOLI SW: P24233"
/codon_start=1
/transl_table=11
/product="nucleoside diphosphate kinase"
/protein_id="AAC75571.1"

promoter
complement(1231..2244)
/gene="yfgA"
/note="synonym: b2516"
complement(1231..2244)
/gene="yfgA"
/function="putative membrane, Not classified"
/note="f337; 100 pct identical to YFGA_ECOLI SW: P27434"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAC75569.1"
/db_xref="GI:1788864"
/translation="MNTTEATHDQNEALTTGARLNRNAREQLGSQLVAERLCLKVSTVRIEEDKAPADLASTPFLRGVRSYARLVHPEBELPGEKAPLEAAVAPQOSFLGKRKRKRDGMVMTFTLVLPVYIGSGAMWMDQKAOEITTMAROSSAEISNSQEGSVPLNITTTDPTTSTPPASVDITANTOTPAYTAAPAVDPQQAIVSSQANVDTAATPAPTAATPDGAAPLPDQAGVATPVADPNLVNPFADCVLEVDATGKLP SGWQRDGNLNTGQAPYKLIKIGAPAAVOIYQGKEVDLSRFIRITQVARTLNAOS PAO"
complement(1288..1316)
/gene="yfgA"
/note="factor Sigma70; predicted +1 start at 2639901"
complement(1360..1389)
/gene="yfgA"
/note="factor Sigma70; predicted +1 start at 2639973"
complement(2529..3683)
/gene="yfgB"
/note="synonym: b2517"
complement(2529..3683)
/gene="yfgB"
/function="orf; Unknown"
/note="f384; 100 pct identical to YFGB_ECOLI SW: P36979"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC75570.1"
/db_xref="GI:1788865"
/translation="MSEQLVTPENVTTKDKINLIDLNRQMRFPKDLGKXPFRADQVKMKTHYCDNDEMDTINKYLRLGKLYEAKRABVVEQSSSGCTIKMALAVDO RAEVTVIPEDDRLTLCVSSQVGCALCKFCSTROQGFNNRLRSEIIGVWRKAYGVAATVYQQRIPITVVMWGEPLNLNNVPAMIMDDGFGLSKRVTLSGVPAIDLGLMIVVALAISLAPNDEIRDIVINKKYNIEFLAAVRYLEKSNNAQGVTLIEVVMDDHNDGEHHAOLAEILKCPCKINILPNMPFGAPYGRSSNSRIDFSKVLMSYGFITVIRKTRGGDDIDACGQLADVIDRTKRTLKRMQGAIDIKAV"
complement(3778..3806)
/note="factor Sigma70; predicted +1 start at 2642391"
complement(3786..3814)
/note="factor Sigma70; predicted +1 start at 2642399"
complement(3833..4264)
/gene="ndk"
/note="synonym: b2518"
complement(3833..4264)
/gene="ndk"
/EC_number="2.7.4.6"
/function="enzyme; Purine ribonucleotide biosynthesis"
/note="f143; 100 pct identical to NDK_ECOLI SW: P24233"
/codon_start=1
/transl_table=11
/product="nucleoside diphosphate kinase"
/protein_id="AAC75571.1"

promoter
complement(1231..2244)
/gene="yfgA"
/note="synonym: b2516"
complement(1231..2244)
/gene="yfgA"
/function="putative membrane, Not classified"
/note="f337; 100 pct identical to YFGA_ECOLI SW: P27434"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AAC75569.1"
/db_xref="GI:1788864"
/translation="MNTTEATHDQNEALTTGARLNRNAREQLGSQLVAERLCLKVSTVRIEEDKAPADLASTPFLRGVRSYARLVHPEBELPGEKAPLEAAVAPQOSFLGKRKRKRDGMVMTFTLVLPVYIGSGAMWMDQKAOEITTMAROSSAEISNSQEGSVPLNITTTDPTTSTPPASVDITANTOTPAYTAAPAVDPQQAIVSSQANVDTAATPAPTAATPDGAAPLPDQAGVATPVADPNLVNPFADCVLEVDATGKLP SGWQRDGNLNTGQAPYKLIKIGAPAAVOIYQGKEVDLSRFIRITQVARTLNAOS PAO"
complement(1288..1316)
/gene="yfgA"
/note="factor Sigma70; predicted +1 start at 2639901"
complement(1360..1389)
/gene="yfgA"
/note="factor Sigma70; predicted +1 start at 2639973"
complement(2529..3683)
/gene="yfgB"
/note="synonym: b2517"
complement(2529..3683)
/gene="yfgB"
/function="orf; Unknown"
/note="f384; 100 pct identical to YFGB_ECOLI SW: P36979"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC75570.1"
/db_xref="GI:1788865"
/translation="MSEQLVTPENVTTKDKINLIDLNRQMRFPKDLGKXPFRADQVKMKTHYCDNDEMDTINKYLRLGKLYEAKRABVVEQSSSGCTIKMALAVDO RAEVTVIPEDDRLTLCVSSQVGCALCKFCSTROQGFNNRLRSEIIGVWRKAYGVAATVYQQRIPITVVMWGEPLNLNNVPAMIMDDGFGLSKRVTLSGVPAIDLGLMIVVALAISLAPNDEIRDIVINKKYNIEFLAAVRYLEKSNNAQGVTLIEVVMDDHNDGEHHAOLAEILKCPCKINILPNMPFGAPYGRSSNSRIDFSKVLMSYGFITVIRKTRGGDDIDACGQLADVIDRTKRTLKRMQGAIDIKAV"
complement(3778..3806)
/note="factor Sigma70; predicted +1 start at 2642391"
complement(3786..3814)
/note="factor Sigma70; predicted +1 start at 2642399"
complement(3833..4264)
/gene="ndk"
/note="synonym: b2518"
complement(3833..4264)
/gene="ndk"
/EC_number="2.7.4.6"
/function="enzyme; Purine ribonucleotide biosynthesis"
/note="f143; 100 pct identical to NDK_ECOLI SW: P24233"
/codon_start=1
/transl_table=11
/product="nucleoside diphosphate kinase"
/protein_id="AAC75571.1"

QY 841 ATGGTACGGTTAAACGGCTGAGACCAACGCTGGAAGATATCATCTCCGATGAGAGTT 900
Db 841 ATGGTACGGTTAAACGGCTGAGACCAACGCTGGAAGATATCATCTCCGATGAGAGTT 900
QY 901 TCGATTATCGGCTGCGTGTGAATGAGCCAGTGAAGCGCTGTTCTACACTCGAGCGTC 960
Db 901 TCGATTATCGGCTGCGTGTGAATGAGCCAGTGAAGCGCTGTTCTACACTCGAGCGTC 960
QY 961 ACCGGCGGCAACAAGAAAGCGGCTCTATGAGAGTGGCTGCGCAAGAAGCGCTGAGAC 1020
Db 961 ACCGGCGGCAACAAGAAAGCGGCTCTATGAGAGTGGCTGCGCAAGAAGCGCTGAGAC 1020
QY 1021 AACACGATATGATGATGACGAGCTGGAAGACGCAATTTGGTGGAAAGCGCAAGTGGAC 1080
Db 1021 AACACGATATGATGATGACGAGCTGGAAGACGCAATTTGGTGGAAAGCGCAAGTGGAC 1080
QY 1081 GAAGCGCGTGAATTGACGTTGACGAGGTTGAAAAATAA 1119
Db 1081 GAAGCGCGTGAATTGACGTTGACGAGGTTGAAAAATAA 1119

RESULT 4
AX393943 1119 bp DNA linear PAT 23-MAR-2002
LOCUS Sequence 3 from Patent WO0212478.
DEFINITION AX393943
VERSION AX393943.1 GI:19701905
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE
1 Boronac, A., Campos, N., Rodriguez-Concepcion, M., Rohmer, M.,
Seeman, M., Valentin, H.E., Venkatesh, T.V. and Venkatesh, M.,
Methyl-d-erythritol phosphate pathway genes
Patent: WO 0212478-A 3 14-FEB-2002;
Monsanto Technology LLC (US)
Location/Qualifiers

FEATURES
source
1. 1119
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
1. 1119
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD29065.1"
/translation="MHWQAPIQRKSTRIVGNVPIGDGAPLAVOSMTNTRTDEAT
VNOIKALERVGADIVRVSPTMDAAEFKIKQOVNPLVADLHFDRILAKAEV
DCIRINGENIERIMVVDKADKNIPIRIGVNASLEKDEKVEPTPALLES
AMRVHDLRLNPFQFVSVKASDVPLAVESYRLAIOIDPLHIGITTEAGRSGLV
KSAIGLILSEIGDPLRVSLADPVEIKVGPDIKSRISRGINFIACPTCSRO
EFVDVITGVNMLEQLELDITTPMDVSIIGCVNPEGELVSTLGVTSNKKSGLYEDSV
RKRLDNDNDIMDLEARIKAKASQLDEARRIDVQVEX"

BASE COUNT 272 a 279 c 317 g 251 t
ORIGIN

Query Match 100.0%; Score 1119; DB 6; Length 1119;
Best Local Similarity 100.0%; Pred. No. 8.6e-266;
Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCATACACGAGCTCCATTCACAGTAAATCAACAGCTATTATTCGTTGGAGATGTG 60
Db 1 ATGCATACACGAGCTCCATTCACAGTAAATCAACAGCTATTATTCGTTGGAGATGTG 60
QY 61 CCGATTGGCGATGGTGTCTCCCATCGCGTACAGTCATGACCAATACGCGTACAGAC 120
Db 61 CCGATTGGCGATGGTGTCTCCCATCGCGTACAGTCATGACCAATACGCGTACAGAC 120
QY 121 GTGAAACAACGCTCATCAATCAAGGCGCTGGAAGCGTGGCGTGAATGTGCTCGT 180

Db 121 GTGAAACAACGCTCATCAATCAAGGCGCTGGAAGCGTGGCGTGAATGTGCTCGT 180
QY 181 GTATCCGTACCGACAGTAGACCGGCGAAGCGTTCAACTCATCAACAGAGTTAAC 240
Db 181 GTATCCGTACCGACAGTAGACCGGCGAAGCGTTCAACTCATCAACAGAGTTAAC 240
QY 241 GTGCCGCTGTGCTGACATCCACTTCGACTATCGGATGGCTGTAAGTAGCGGAATTC 300
Db 241 GTGCCGCTGTGCTGACATCCACTTCGACTATCGGATGGCTGTAAGTAGCGGAATTC 300
QY 301 GGCGTGATTTGTCGCGTATTAACCTGCGCAATATCGTAATGAAAGAGGTATTCACATG 360
Db 301 GGCGTGATTTGTCGCGTATTAACCTGCGCAATATCGTAATGAAAGAGGTATTCACATG 360
QY 361 GTGTTGACTGTGCGCGCATTAATAATTCGATTCGATTGGCGTTAACCGCGATGCG 420
Db 361 GTGTTGACTGTGCGCGCATTAATAATTCGATTCGATTGGCGTTAACCGCGATGCG 420
QY 421 CTGGAATAAAGATCTGCAAGAAAGTATGGGGAACGACCGCGGCGTTGCTGGAATCT 480
Db 421 CTGGAATAAAGATCTGCAAGAAAGTATGGGGAACGACCGCGGCGTTGCTGGAATCT 480
QY 481 GCCATCGTCATGTTGATCATCTCGATCCGCTGAACCTTGATCGTCAAAAGTCACGCTG 540
Db 481 GCCATCGTCATGTTGATCATCTCGATCCGCTGAACCTTGATCGTCAAAAGTCACGCTG 540
QY 541 AAAGCGTCTGACGCTCTTCTCGCTGTTAGTCTTATGCTTTGCTGGCAAAACAGATGAT 600
Db 541 AAAGCGTCTGACGCTCTTCTCGCTGTTAGTCTTATGCTTTGCTGGCAAAACAGATGAT 600
QY 601 CAGCCGTCGATCTGGGATCAACCGAAGCGGCGGCGCGGCGGCGAGTAAATTC 660
Db 601 CAGCCGTCGATCTGGGATCAACCGAAGCGGCGGCGGCGGCGGCGAGTAAATTC 660
QY 661 GCATTGTTAGTGTGCTGCTGTCTGATGAAGGATCGGCGACACGCTGCGTATCGCTG 720
Db 661 GCATTGTTAGTGTGCTGCTGTCTGATGAAGGATCGGCGACACGCTGCGTATCGCTG 720
QY 721 GCGCGCGATCCGCTGCAAGAGATCAAGTGGCTTGCATATTTGAAATCGCTGCGTATC 780
Db 721 GCGCGCGATCCGCTGCAAGAGATCAAGTGGCTTGCATATTTGAAATCGCTGCGTATC 780
QY 781 CGTTCCGAGGATCACTTCATCGGCTCCGACCTGTTCCGCTGAGGAATTTGATGTT 840
Db 781 CGTTCCGAGGATCACTTCATCGGCTCCGACCTGTTCCGCTGAGGAATTTGATGTT 840
QY 841 ATCGTACGTTTAAACGCTGAGCAACGCTGGAAGATATCATCTCCGATGAGAGTT 900
Db 841 ATCGTACGTTTAAACGCTGAGCAACGCTGGAAGATATCATCTCCGATGAGAGTT 900
QY 901 TCGATTATCGGCTGCGTGTGAATGAGCCAGTGAAGCGCTGTTCTACACTCGAGCGTC 960
Db 901 TCGATTATCGGCTGCGTGTGAATGAGCCAGTGAAGCGCTGTTCTACACTCGAGCGTC 960
QY 961 ACCGGCGGCAACAAGAAAGCGGCTCTATGAGAGTGGCTGCGCAAGAAGCGCTGAGAC 1020
Db 961 ACCGGCGGCAACAAGAAAGCGGCTCTATGAGAGTGGCTGCGCAAGAAGCGCTGAGAC 1020
QY 1021 AACACGATATGATGATGACGAGCTGGAAGACGCAATTTGGTGGAAAGCGCAAGTGGAC 1080
Db 1021 AACACGATATGATGATGACGAGCTGGAAGACGCAATTTGGTGGAAAGCGCAAGTGGAC 1080
QY 1081 GAAGCGCGTGAATTGACGTTGACGAGGTTGAAAAATAA 1119
Db 1081 GAAGCGCGTGAATTGACGTTGACGAGGTTGAAAAATAA 1119

RESULT 5
ECGCP ECGCP 1697 bp DNA linear BCT 08-DEC-1993
LOCUS E.coli gcpE gene.
DEFINITION X64451 543432

QY 421 CTGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCGAGGCGTTCTGGAACTT 480
 DB 421 CTGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCGAGGCGTTCTGGAACTT 480
 QY 481 GCCATGGGTATGTTGATTCATCTCGATCGCTGAACCTTCATCTCAAGTACAGCTG 540
 DB 481 GCCATGGGTATGTTGATTCATCTCGATCGCTGAACCTTCATCTCAAGTACAGCTG 540
 QY 541 AAGCGCTGACGCTCTCTCGCTGTGAGCTTATGTTGCTGGCAAAACAGATCAT 600
 DB 541 AAGCGCTGACGCTCTCTCGCTGTGAGCTTATGTTGCTGGCAAAACAGATCAT 600
 QY 601 CAGCGCTGATCTGGGAGATCACGAAAGCCGCTGCGCGACGCGGAGTAAATCC 660
 DB 601 CAGCGCTGATCTGGGAGATCACGAAAGCCGCTGCGCGACGCGGAGTAAATCC 660
 QY 661 GCCATTGGTTAAGTCTGCTGCTGTCTGAAGGATCGGCGACACAGCTGCGCTATCGCTG 720
 DB 661 GCCATTGGTTAAGTCTGCTGCTGTCTGAAGGATCGGCGACACAGCTGCGCTATCGCTG 720
 QY 721 GCGGCGCATCCGCTGCAAGAGATCAAGTCCGTTTCGATATTTTGAATGCGCTGATC 780
 DB 721 GCGGCGCATCCGCTGCAAGAGATCAAGTCCGTTTCGATATTTTGAATGCGCTGATC 780
 QY 781 CGTTCGAGAGATCAACTTCATCGCTGCGACCTGTTGCGCTCAGAAATTTGATGT 840
 DB 781 CGTTCGAGAGATCAACTTCATCGCTGCGACCTGTTGCGCTCAGAAATTTGATGT 840
 QY 841 ATGGGTTCGGTTAACGCGCTGAGCAAGCGCTGGAAGATTCATCACTCCGATGACGTT 900
 DB 841 ATGGGTTCGGTTAACGCGCTGAGCAAGCGCTGGAAGATTCATCACTCCGATGACGTT 900
 QY 901 TCGATTTATCGGCTGCGTGAATGGCCAGGTGAGGCGCGGTTTCTACCTCGGCGTC 960
 DB 901 TCGATTTATCGGCTGCGTGAATGGCCAGGTGAGGCGCGGTTTCTACCTCGGCGTC 960
 QY 961 ACCGCGCGCAACAGAAAAGCGGCTCTATGAAAGTGGCGTGGCAAGACCGCTGGAC 1020
 DB 961 ACCGCGCGCAACAGAAAAGCGGCTCTATGAAAGTGGCGTGGCAAGACCGCTGGAC 1020
 QY 1021 AACACGATATGATCGACCAAGCTGGAAGCAAGCTTGTGCGAAAGCGATCAGCTGAC 1080
 DB 1021 AACACGATATGATCGACCAAGCTGGAAGCAAGCTTGTGCGAAAGCGATCAGCTGAC 1080
 QY 1081 GAAGCGCGTGAATGAGCTTCAGCAGGTTGAAAAATTA 1119
 DB 1081 GAAGCGCGTGAATGAGCTTCAGCAGGTTGAAAAATTA 1119

RESULT 3
 AX050487 1119 bp DNA linear PART 12-JAN-2001
 LOCUS AX050487
 DEFINITION Sequence 1 from Patent WO0072022.
 ACCESSION AX050487
 VERSION AX050487.1 GI:12226691
 KEYWORDS
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 1
 Jomaa, H.
 Use of genes of the deoxy-d-xylose phosphate biosynthetic pathway
 for altering the concentration of isoprenoid
 Patent: WO 0072022-A 1 30-NOV-2000;
 Jomaa, Hassen (DE)
 location/Qualifiers
 1..1119
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /db_xref="taxon:562"
 1..1119
 /note="unnamed protein product"

/codon_start=1
 /transl_table=11
 /protein_id="CAC1801.1"
 /db_xref="GI:12226692"
 /translation="MHNQAPIQRRKSTRIVGVNPVIGDAPLAVGSMTRTTTDEAT
 VNOIKALERVADIVRVSPTMDAAEFLLIQVNVPLVADHDFYRLALKVAEYGV
 DCLRIHPNGINBERIRMVDCARDKNIIRIGVNAISLEKDIQERYGEPDALLS
 ARMHVDHLNFDQPKVSKASDVPLAIESYRLAKQIDOPHLGITEAGASGAV
 KSAIGGLLSGIGDITLRLSLAADVERIKVGFLLKSLIRSRGINFACPTGSRQ
 EFDVIGTVALSQRLEDITTPMDVSIIGCVNNGPGLALYSLTGVTGSKKSGLYEDGV
 RKDRLDNMIMDLEAIRAKASQLEBARRIDVQVER"

BASE COUNT 272 a 279 c 317 g 251 t

ORIGIN

Query Match 100.0%; Score 1119; DB 6; Length 1119;
 Best Local Similarity 100.0%; Pred. No. 8, 6e-266;
 Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCATACAGGCTCCAAATTCACGTAGAAAATCAACAGTATTTACGTTGGAATGTT 60
 DB 1 ATGCATACAGGCTCCAAATTCACGTAGAAAATCAACAGTATTTACGTTGGAATGTT 60
 QY 61 CCGATTGGCGATGCTCTCCATCGCGCTACAGTCCATACCAATACCGGTACAGAC 120
 DB 61 CCGATTGGCGATGCTCTCCATCGCGCTACAGTCCATACCAATACCGGTACAGAC 120
 QY 121 GTGAGCAACGGTCAATCAATCAAGGCGCTGGAAAGCGTTGGCGCTGATATCGTCCGT 180
 DB 121 GTGAGCAACGGTCAATCAATCAAGGCGCTGGAAAGCGTTGGCGCTGATATCGTCCGT 180
 QY 181 GTATCCGTCACGATGAGACGCGGAGAGAGCTTCAACTCATCAACAGCAGGTTAAC 240
 DB 181 GTATCCGTCACGATGAGACGCGGAGAGAGCTTCAACTCATCAACAGCAGGTTAAC 240
 QY 241 GTGCGCTGCTGAGTCACTTCATCGATTCGATTCGCTGAAAGTACGGAATAC 300
 DB 241 GTGCGCTGCTGAGTCACTTCATCGATTCGATTCGCTGAAAGTACGGAATAC 300
 QY 301 GGGTGTGATTCGTGGGTATTAACCTGGCAATATGGTAAGAGGATATTCGATG 360
 DB 301 GGGTGTGATTCGTGGGTATTAACCTGGCAATATGGTAAGAGGATATTCGATG 360
 QY 361 GTGATGACGTGCGGCGCATAAAACATTCGATCCGATTTAGCGTTAACGCGGATG 420
 DB 361 GTGATGACGTGCGGCGCATAAAACATTCGATCCGATTTAGCGTTAACGCGGATG 420
 QY 421 CTGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCGAGGCGTTCTGGAACTT 480
 DB 421 CTGAAAAAGATCTGCAAGAAAAGTATGGCGAACCGACGCCGCGAGGCGTTCTGGAACTT 480
 QY 481 GCCATGGGTATGTTGATTCATCTCGATCGCTGAACCTTCATCTCAAGTACAGCTG 540
 DB 481 GCCATGGGTATGTTGATTCATCTCGATCGCTGAACCTTCATCTCAAGTACAGCTG 540
 QY 541 AAGCGCTGACGCTCTCTCGCTGTGAGCTTATGTTGCTGGCAAAACAGATCAT 600
 DB 541 AAGCGCTGACGCTCTCTCGCTGTGAGCTTATGTTGCTGGCAAAACAGATCAT 600
 QY 601 CAGCGCTGATCTGGGAGATCACGAAAGCCGCTGCGCGACGCGGAGTAAATCC 660
 DB 601 CAGCGCTGATCTGGGAGATCACGAAAGCCGCTGCGCGACGCGGAGTAAATCC 660
 QY 661 GCCATTGGTTAAGTCTGCTGCTGTCTGAAGGATCGGCGACACAGCTGCGCTATCGCTG 720
 DB 661 GCCATTGGTTAAGTCTGCTGCTGTCTGAAGGATCGGCGACACAGCTGCGCTATCGCTG 720
 QY 721 GCGGCGCATCCGCTGCAAGAGATCAAGTCCGTTTCGATATTTTGAATGCGCTGATC 780
 DB 721 GCGGCGCATCCGCTGCAAGAGATCAAGTCCGTTTCGATATTTTGAATGCGCTGATC 780
 QY 781 CGTTCGAGAGATCAACTTCATCGCTGCGACCTGTTGCGCTCAGAAATTTGATGT 840
 DB 781 CGTTCGAGAGATCAACTTCATCGCTGCGACCTGTTGCGCTCAGAAATTTGATGT 840

GPC AG GENOME PHARMACEUTICALS (DE)
 FEATURES
 Location/Qualifiers
 1. 1119
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /db_xref="taxon:562"
 BASE COUNT 272 a 279 c 317 g 251 t
 ORIGIN
 Query Match 100.0%; Score 1119; DB 6; Length 1119;
 Best Local Similarity 100.0%; Pred. No. 8.6e-266;
 Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCATACCAAGGCTCCATTCACAGTAAATCAACAGGATTTACGTTGGGAATGTG 60
 DB 1 ATGCATACCAAGGCTCCATTCACAGTAAATCAACAGGATTTACGTTGGGAATGTG 60

QY 61 CCAGTTGGCGATGGTGTCTCCCATGCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
 DB 61 CCAGTTGGCGATGGTGTCTCCCATGCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

QY 121 GTCCAGCAACAGGCTCAATCAATCAAGCGCTGGAGCGGTTGGCGCTGATATGTCGCT 180
 DB 121 GTCCAGCAACAGGCTCAATCAATCAAGCGCTGGAGCGGTTGGCGCTGATATGTCGCT 180

QY 181 GTATCCGATCCAGATGAGACGCGCGCAAGCGTTCAACTCATCAACAGCAGATTAC 240
 DB 181 GTATCCGATCCAGATGAGACGCGCGCAAGCGTTCAACTCATCAACAGCAGATTAC 240

QY 241 GTGCGCGTGGTGGTGATCATCTGATGACATGATGCTTGGCGTGAAGTAGCGGAATAC 300
 DB 241 GTGCGCGTGGTGGTGATCATCTGATGACATGATGCTTGGCGTGAAGTAGCGGAATAC 300

QY 301 GCGCTGCATTTCTGCGCTATTAACTCGCAATTCGATTCGATTAAGAGCGATTTCGATG 360
 DB 301 GCGCTGCATTTCTGCGCTATTAACTCGCAATTCGATTCGATTAAGAGCGATTTCGATG 360

QY 361 GTGCTTGAATGTCGCGCGATTAACATTCCTGATTCGCTTGAAGTAGCGCGATTCG 420
 DB 361 GTGCTTGAATGTCGCGCGATTAACATTCCTGATTCGCTTGAAGTAGCGCGATTCG 420

QY 421 CTGGAATAAGATCTGCAAGAAATATGAGCAACGACGCGCGGCTTGCTGGAATCT 480
 DB 421 CTGGAATAAGATCTGCAAGAAATATGAGCAACGACGCGCGGCTTGCTGGAATCT 480

QY 481 GCCATGCGTCAATGTTGATCATCTGATCGCCTGAATCTCGATCAAGTCAAGCTG 540
 DB 481 GCCATGCGTCAATGTTGATCATCTGATCGCCTGAATCTCGATCAAGTCAAGCTG 540

QY 541 AAAGCGTCTGACGTTCTCTCTGCTGTTGATGTTATGTTTGGCGAAACAGATCAT 600
 DB 541 AAAGCGTCTGACGTTCTCTCTGCTGTTGATGTTATGTTTGGCGAAACAGATCAT 600

QY 601 CAGCGGTGATCTGGGGATCACGGAAGCGGTGGTGCGGCGAGCGGGGAGTAAATCC 660
 DB 601 CAGCGGTGATCTGGGGATCACGGAAGCGGTGGTGCGGCGAGCGGGGAGTAAATCC 660

QY 661 GCCATTGGTTAGTGTCTGCTGCTGTTGAAGCATTCGCGCAACAGCTGCGGTATCCG 720
 DB 661 GCCATTGGTTAGTGTCTGCTGCTGTTGAAGCATTCGCGCAACAGCTGCGGTATCCG 720

QY 721 GCGGCGCATTCGCTGCGTGAAGAGATCAAAAGTCGTTTCGATATTTTGAATTCGTCG 780
 DB 721 GCGGCGCATTCGCTGCGTGAAGAGATCAAAAGTCGTTTCGATATTTTGAATTCGTCG 780

QY 781 CGTTGCGAGGAGATCAATTCATCGCTGCGCGACCTGTTGCGTCAAGAAATTTGAATG 840
 DB 781 CGTTGCGAGGAGATCAATTCATCGCTGCGCGACCTGTTGCGTCAAGAAATTTGAATG 840

QY 841 ATCCGATCGTTAAACGCGCTGAGCAACGCTGGAAGATATCATCACTCCGATGAGCGTT 900
 DB 841 ATCCGATCGTTAAACGCGCTGAGCAACGCTGGAAGATATCATCACTCCGATGAGCGTT 900

QY 901 TCGATTATCGGCTCGCTGGTGAATGCGCCAGGTGAGCGCGTGGTTCTACATCGGCGTC 960
 DB 901 TCGATTATCGGCTCGCTGGTGAATGCGCCAGGTGAGCGCGTGGTTCTACATCGGCGTC 960

QY 961 ACCGCGGCAACAAGAAACCGGCTCTATGAAAGATGGCGTGGCAAGACCGCTGAGAC 1020
 DB 961 ACCGCGGCAACAAGAAACCGGCTCTATGAAAGATGGCGTGGCAAGACCGCTGAGAC 1020

QY 1021 AACACGATATGATGACCACTGAGACACCGCATTCGTTCCGAAGCCGATGAGTGAAC 1080
 DB 1021 AACACGATATGATGACCACTGAGACACCGCATTCGTTCCGAAGCCGATGAGTGAAC 1080

QY 1081 GAAGCGGTGCAATGACGTTGAGTGAATAA 1119
 DB 1081 GAAGCGGTGCAATGACGTTGAGTGAATAA 1119

RESULT 2
 AX038912 1119 bp DNA linear PAT 16-NOV-2000
 LOCUS
 DEFINITION Sequence 29 from Patent W00061793.
 AX038912
 ACCESSION
 VERSION AX038912.1 GI:11228221
 KEYWORDS
 SOURCE
 ORGANISM
 Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE
 1
 Lofrer, H. and Jacob, A.
 Novel method for identifying antibacterial compounds
 Patent: WO 0061793-A 29 19-OCT-2000;
 GPC BIOTECH AG (DE); LOFRER HANNES (DE); JACOBI ALEXANDER (DE)

FEATURES
 source
 1. 1119
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /db_xref="taxon:562"
 BASE COUNT 272 a 279 c 317 g 251 t
 ORIGIN
 Query Match 100.0%; Score 1119; DB 6; Length 1119;
 Best Local Similarity 100.0%; Pred. No. 8.6e-266;
 Matches 1119; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCATACCAAGGCTCCATTCACAGTAAATCAACAGTATTTACGTTGGGAATGTG 60
 DB 1 ATGCATACCAAGGCTCCATTCACAGTAAATCAACAGTATTTACGTTGGGAATGTG 60

QY 61 CCAGTTGGCGATGGTGTCTCCCATGCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120
 DB 61 CCAGTTGGCGATGGTGTCTCCCATGCGCGTACAGTCCATGACCAATACGCGTACGACAGAC 120

QY 121 GTCCAGCAACAGGCTCAATCAATCAAGCGCTGGAGCGGTTGGCGCTGATATGTCGCT 180
 DB 121 GTCCAGCAACAGGCTCAATCAATCAAGCGCTGGAGCGGTTGGCGCTGATATGTCGCT 180

QY 181 GTATCCGATCCAGATGAGACGCGCGCAAGCGTTCAACTCATCAACAGCAGATTAC 240
 DB 181 GTATCCGATCCAGATGAGACGCGCGCAAGCGTTCAACTCATCAACAGCAGATTAC 240

QY 241 GTGCGCGTGGTGGTGATCATCTGATGACATGATGCTTGGCGTGAAGTAGCGGAATAC 300
 DB 241 GTGCGCGTGGTGGTGATCATCTGATGACATGATGCTTGGCGTGAAGTAGCGGAATAC 300

QY 301 GCGCTGCATTTCTGCGCTATTAACTCGCAATTCGATTCGATTAAGAGCGATTTCGATG 360
 DB 301 GCGCTGCATTTCTGCGCTATTAACTCGCAATTCGATTCGATTAAGAGCGATTTCGATG 360

QY 361 GTGCTTGAATGTCGCGCGATTAACATTCCTGATTCGCTTGAAGTAGCGCGATTCG 420
 DB 361 GTGCTTGAATGTCGCGCGATTAACATTCCTGATTCGCTTGAAGTAGCGCGATTCG 420

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 23, 2003, 13:04:06 ; Search time 4350 Seconds

(without alignments)
10523.649 Million cell updates/sec

Title: US-09-921-992-3

Perfect score: 1119
Sequence: 1 atgcataaccagagctcccaat.....ttccagcaggttgaaataaa 1119Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapept 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*

1: gb_ba:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pac:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_scs:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pac:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_scs:*

28: em_un:*

29: em_vl:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pln:*

35: em_hcg_rod:*

36: em_hcg_mam:*

37: em_hcg_vrt:*

38: em_hcg_hum:*

39: em_hcg_mus:*

40: em_hcg_rod:*

41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1119	100.0	1119	6	AX036302 Sequence
2	1119	100.0	1119	6	AX038912 Sequence
3	1119	100.0	1119	6	AX050487 Sequence
4	1119	100.0	1119	6	AX393943 Sequence
5	1119	100.0	1697	1	ECGCEP
6	1119	100.0	13176	1	AE000338 Escherich
7	1098.4	98.2	10225	1	AE015271 Shigella
8	1098.4	98.2	290380	1	AE016987 Shigella
9	1098.2	98.1	300099	1	AE016764 Escherich
10	1093.4	97.7	1119	1	AY033515 Escherich
11	1093.4	97.7	11521	1	AE005481 Escherich
12	1093.4	97.7	296827	1	AP002561 Escherich
13	1010	90.3	1010	6	AR027993 Sequence
14	1010	90.3	1010	6	AR208328 Sequence
15	1010	90.3	1010	6	BD084908 Methods f
16	887	79.3	23647	1	AE008814 Salmonell
17	885.4	79.1	145050	1	AL627275 Salmonell
18	885.4	79.1	300431	1	AE016835 Salmonell
19	885.4	79.1	4775	1	D90881 E.coli geno
20	690	61.7	16214	1	D90880 E.coli geno
21	685.4	61.3	65821	2	AC020826 Mus muscu
22	665	59.4	11824	1	AE004161 Vibrio ch
23	658.4	58.8	301235	1	AP005075 Vibrio pa
24	632.4	56.5	301442	1	AE016798 Vibrio pa
25	631.6	56.4	1345	1	PSU67933 Providencia
26	631.6	56.4	1345	6	AR207991 Sequence
27	631.6	56.4	1345	6	AR208326 Sequence
28	631.6	56.4	1345	6	BD084906 Methods f
29	622	55.6	10507	1	U32721 Haemophilus
30	622	55.6	110000	6	AR274513_03 Continuation (4 of
31	620.6	55.5	11161	1	AE013738 Yersinia
32	620.6	55.5	204050	1	AJ141454 Yersinia
33	619.8	55.4	1079	6	AR027992 Sequence
34	619.8	55.4	1079	6	AR208327 Sequence
35	619.8	55.4	1079	6	BD084907 Methods f
36	603.6	53.9	11268	1	AE006237 Pasteurel
37	566	50.6	10607	1	AE015769 Shewanell
38	534.6	47.8	886	6	AR027995 Sequence
39	534.6	47.8	886	6	AR208330 Sequence
40	534.6	47.8	886	6	BD084910 Methods f
41	533	47.6	300861	1	AE016777 Pseudomon
42	525	46.9	311249	1	AE016860 Pseudomon
43	515.4	46.1	17148	1	AE004798 Pseudomon
44	403.6	36.1	110000	6	BD061520_3 Continuation (4 of
45	403.6	36.1	347550	1	AP001118 Buchnera

ALIGNMENTS

RESULT 1
LOCUS AX036302 1119 bp DNA
DEFINITION Sequence 29 from Patent EP1043403.
ACCESSION AX036302
VERSION AX036302.1 GI:11225912
KEYWORDS
SOURCE
ORGANISM Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Novel method for identifying antibacterial compounds
Patent: EP 1043403-A 29 11-OCT-2000;
JOURNAL

OY 307 ArglySGlyTyrValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGly 326
 |||||.....
 DB 125 CGGAAGGGTGGCTGGCCGCCCACTCATCAAGCTGTAATTTTACAGATTAATGGGA 184
 OY 327 PheIleIleGlnGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGly 346
 |||||.....
 DB 185 TTTATCATTTAGCAGATATATTAACCGAATGGCGAAGCTCAGACATCCATTGGAAGA 244
 OY 347 AspLeuLeuTyrAlaIleGlnArgValLeuLysLeuSerValProAsnLeuTyrValTyr 366

 DB 245 AACCTTTATACGCCCATCCAGAGGATATGAACTTTCAGTTCAATTTATATATGCTGG 304
 OY 367 LeuCysMetPheTyrCysPhePheHisLeuTyrLeuAsnIleLeuAlaGluLeuLys 386
 |||||.....
 DB 305 CTTGTCATGTTCTACAGCTCTTTCATCTTGGTTAAATATACCTCGGAAGTTCTGGCA 364
 OY 387 PheGlyAspArgGlnPheTyrLysAspTyrPheAspAlaLysSerValGlyAspTyr 406
 |||||.....
 DB 365 TTTGGGATCGTCACTTATATAAGATTTGGTGCAACGCAAAACAATTGATGATGCTG 424
 OY 407 ArgMetTrpAsnMetProValHisLysTyrMetValArgHisIleTyrPheProCysLeu 426
 |||||.....
 DB 425 AGACTTTGAATATGCTGCTGATAGTGCATGCTGCTGCACATCTATTTCCATGCTTA 484
 OY 427 ArgSerLysIleProLysThrLeuAlaIleIleIleAlaPheLeuValSerAla 444
 |||||.....
 DB 485 AGAAATGACATACCTAAGGAGGATGCAATGATGATCTTTCTTATATCTGCT 538
 RESULT 12
 B316561
 LOCUS B316561 606 bp mRNA linear EST 09-APR-2002
 DEFINITION B316561 y. Ogihara unpublished cDNA library, Wh_yf Trilicium
 accession cDNA clone whzf23117 5', mRNA sequence.
 B316561
 VERSION B316561.1 GI:20121748
 KEYWORDS EST
 SOURCE Trilicium aestivum (bread wheat)
 ORGANISM Trilicium aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 Triliceae; Trilicium.
 1 (bases 1 to 606)
 Ogihara, Y. and Mural, K.
 Unpublished
 Expressed genes in Trilicium aestivum
 COMMENT
 Center for Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshn1genes.nlg.ac.jp.
 Location/Qualifiers
 1..606
 /organism="Trilicium aestivum"
 /mol_type="mRNA"
 /cultivar="Chinese Spring"
 /db_xref="taxon:4565"
 /clone="whzf23117"
 /tissue_type="spikelet at early flowering"
 /dev_stage="Peekes" scale 6"
 /clone_1lb="Y. Ogihara unpublished cDNA library, Wh_yf"
 BASE COUNT 170 a 113 c 125 g 198 t
 ORIGIN
 Alignment Scores:
 Pired. No.: 1,22e-63 Length: 606
 Score: 762.50 Matches: 136
 Percent Similarity: 79.21% Conservative: 24
 Best Local Similarity: 67.33% Mismatches: 35
 Query Match: 27.52% Indels: 7
 DB: 12 Gaps: 1

US-09-623-514a-2 (1-520) x B316561 (1-606)
 OY 238 GlyValThrLeuMetLeuThrCysIleValIleTrpLeuLysLeuValSerTyrAlaHis 257
 |||||.....
 DB 1 GGTTTTGTGATGATGTTATATGCTTCATGCTTCATGCTTCATGCTTCATGCTTCAT 60
 OY 258 ThrSerTyrAspIleArgSerLeuAla-----AsnAlaAlaAsp 270
 |||||.....
 DB 61 ACAACCATGACATATAGGCAATTAACCAATAGTGTAGAGGTGATATATGACACCAC 120
 OY 271 LysAlaAsnProGluValSerTyrTyrValSerLeuLysSerLeuAlaTyrPheMetVal 290

 DB 121 ACAGCTGACATGATGATATTTACAGCTCCAACTTTGAGGACCTATATATCTTCAATG 180
 OY 291 AlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArgLysGlyTyr 310
 |||||.....
 DB 181 GCTCCACACCTGCTGATACGCCAAGTTATCTCGAAGTGAATAATGTTGAAAAAGGCTG 240
 OY 311 ValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleGln 330

 DB 241 CGATTCGCAATATTTCTGCACTGATATTTACAGGATATCAAGTTTCATATTTAG 300
 OY 331 GlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeuTyr 350
 |||||.....
 DB 301 CAGTACATTAATCCAACTCGTGAACCTCCAACTCCATTCGAAAGCGGACTTCTGAAT 360
 OY 351 AlaIleGlnArgValLeuLysLeuSerValProAsnLeuTyrValIleTrpLeuCysMetPhe 370
 |||||.....
 DB 361 GCTGTAGACACTGTTGTGAGGCTTCATACCAAAAGTTTACTTATGCTTTCAGATGTC 420
 OY 371 TyrCysPhePheHisLeuTyrLeuAsnIleLeuAlaGluLeuLysPheGlyAspArg 390
 |||||.....
 DB 421 TATGCTTTTTCATCTCGTGTAAATATPCTGAGATCTTCTGCTTTCGTTGCGTACCG 480
 OY 391 GluPheTyrLysAspTyrPheAspAlaLysSerValGlyAspTyrTrpArgMetTrpAsn 410
 |||||.....
 DB 481 GAATTCACAAACACGCTGAGATGCAAAACAATGATGAGATGCGGAAAAATGGAAC 540
 OY 411 MetProValHisLysTyrMetValArgHisIleTyrPheProCysLeuSerLysIle 430
 |||||.....
 DB 541 ATGCCCGTCATTAATGATGTTGCTTCATATATATTTCTTGCATGCAAGTGGTATTA 600
 OY 431 ProLys 432
 |||||.....
 DB 601 TCGAAG 606
 RESULT 13
 AW349274/c
 LOCUS AW349274 855 bp mRNA linear EST 04-OCT-2000
 DEFINITION GM210004B21H12 Gm-r1021 Glycine max cDNA clone Gm-r1021-1536 3',
 mRNA sequence.
 AW349274
 ACCESSION AW349274.1 GI:6846984
 VERSION EST
 KEYWORDS Glycine max (soybean)
 SOURCE Glycine max
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 855)
 Vodka, L., Kelm, P., Shoemaker, R., Retzel, E., Khanna, A., Corryell, V.,
 Expelling, D., Raph, C., Shoop, E., Pardinas, J., Liu, L., and Lewin, H.
 A Functional Genomics Program for Soybean (NSF 9872565)
 COMMENT
 Other-ESTs: A1441040
 Contact: Vodka, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA

```

VERSION      BJ322102.1  GI:20125036
KEYWORDS     EST
SOURCE       Triticum aestivum (bread wheat)
ORGANISM     Triticum aestivum
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
              ; Triticeae; Triticum
              1 (bases 1 to 729)
REFERENCE    Ogihara, Y. and Murai, K.
AUTHORS      Ogihara, Y. and Murai, K.
TITLE        Expressed genes in Triticum aestivum
JOURNAL      Unpublished
COMMENT      Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              1111 Yatae, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tshin@genes.nig.ac.jp.

FEATURES
  source
    1..729
    /organism="Triticum aestivum"
    /mol_type="mRNA"
    /cultivar="Chinese Spring"
    /db_xref="taxon:4565"
    /clone="whyf23117"
    /tissue_type="spikelet at early flowering"
    /dev_stage="peaks" scale 6"
    /clone_id="Y. Ogihara unpublished cDNA library, whyf"

BASE COUNT   237 a 146 c 153 g 193 t

ORIGIN
Alignment Scores:
Pred. No.:    2,53e-70      Length:      729
Score:        832.50       Matches:     147
Percent Similarity: 84.13%  Conservative: 28
Best Local Similarity: 70.67%  Mismatches: 30
Query Match:  30.04%       Indels:      3
DB:           12          Gaps:         1

US-09-623-514A-2 (1-520) x BJ322102 (1-729)
OY 313 ArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleIleGlu-GlnTy 332
DB 729 GGTCAAAATTATTCGCACTGTATATTACAGGTATTCAGAGTTTCATTATTCAGACAGTA 670
OY 332 rIleAsnProIleValAlaGAsnSer-LysHisProLeuLysGlyAspLeuLeuTyralaI 352
DB 669 CATAAATTCATTCGTCGCACTCTCCACATCCATTCGAAAGCGGACTTCGAAATGCTG 610
OY 352 leGluArgValLeuLysLeuSerValProAsnLeuTyValTrpLeuLysMetPheTyrc 372
DB 609 TAGAGACTGTTCTGAGCCTCATACCAAAATGTTACTTATGCTTGACATGTTCTATW 550
OY 372 yspPheHisLeuTrpLeuAsnIleLeuAlaGluLeuLysGlyAspAlaGlu 392
DB 549 GCTTTTCCATCTCGTGGTTAAATATCTGCGAATTCCTGTTGCTGCGTGCACCGTGAAT 490
OY 392 heTyLysAspTrpTrpAsnAlaLysSerValGlyAspTyTrpArgMetTrpAsnMetP 412
DB 489 TCTACAAAGACTGGTGGAAATGCAAAACATTTGATGAGTATTTGGGAAATGCAACATGC 430
OY 412 rovalHisLysTrpMetValArgHisIleTyPheProCysLeuArgSerLysIleProL 432
DB 429 CCGTCATAAATGATGTTCTGCTCATATATATTTTCCTTGATGCGAAGTGGTATATGGA 370
OY 432 ysrHLeuAlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLeuCysIleA 452
DB 369 AGGAAGTGTCTGTTTGTATCATTTTGTATCTGCGCTGCATGAGCTTGTGTTG 310
OY 452 laValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetHeGlnValProL 472
DB 309 CTGTCCCTCGCGGAATTCGTAAGTTCTGCGCATTCCTAGGAGATCATGCTGCAGATCCCC 250

```

```

OY 472 euValPheIleThrAsnTyLeuGlnGluAqPheGlySerTherVal---GlyAsnMetI 491
DB 249 TTATCATTGTGACATCATGCGTCAAAAGCAATTCAGGATGAGTATGCGCCGCAACATGA 190
OY 491 lePheTrpPheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTyTrpHisA 511
DB 189 TATTTGGTCTTTTGTGCACTACGACGACCATATGCTGCTCTGCTGACTACCATG 130
OY 511 spleuMetAsnArgLysGly 517
DB 129 ATGTGATGACAGCATGTGGC 110

RESULT 11
BI422212      539 bp  mRNA  linear  EST 16-AUG-2001
LOCUS        EST532878 tomato callus, TAMU Lycopersicon esculentum cDNA clone
DEFINITION   c18C69M10 5' end, mRNA sequence.
ACCESSION    BI422212.1  GI:15196417
VERSION      BI422212.1
KEYWORDS     EST.
SOURCE       Lycopersicon esculentum (tomato)
ORGANISM     Lycopersicon esculentum
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
              asterids; lamids; Solanales; Solanaceae; Solanum; Lycopersicon.
              1 (bases 1 to 539)
REFERENCE    Alcala, J., Vrebalov, J., White, R., Matern, A.L., Vision, T., Holt, I.E.,
              Liang, F., Upton, J., Craven, M.B., Bowman, C.L., Ahn, S., Rönning,
              C.M., Fraser, C.M., Martin, G.B., Tanksley, S.D. and Giovannoni, J.
              Generation of ESTs from tomato callus tissue
              Unpublished
              Contact: CUGI
              Clemson University Genomics Institute
              Clemson University
              100 Jordan Hall, Clemson, SC 29634, USA
              Email: http://www.genome.clemson.edu/orders/index.html.

FEATURES
  source
    1..539
    /organism="Lycopersicon esculentum"
    /mol_type="mRNA"
    /cultivar="TA496"
    /db_xref="taxon:4081"
    /clone="c18C69M10"
    /tissue_type="callus"
    /dev_stage="25-40 days old"
    /lab_host="X11-Blue MRF"
    /clone_id="tomato callus, TAMU"
    /note="Vector: pBluescript SK(-); Site 1: EcoRI, Site 2:
    XhoI; supplier: Giovannoni laboratory; c18C - Cotyledons
    of seedlings 7-10 days post-germination were excised, cut
    at both ends and placed on MS medium with no selection.
    Mixed callus was harvested at 25 and 40 days and included
    undifferentiated masses. Tomato callus EST library"

BASE COUNT   137 a 102 c 120 g 180 t

ORIGIN
Alignment Scores:
Pred. No.:    5.91e-64      Length:      539
Score:        765.00       Matches:     136
Percent Similarity: 87.08%  Conservative: 19
Best Local Similarity: 76.40%  Mismatches: 23
Query Match:  27.61%       Indels:      0
DB:           12          Gaps:         0

US-09-623-514A-2 (1-520) x BI422212 (1-539)
OY 267 AsnaIaAlaAspLysAlaAsnProGluValSerTyTrpValSerLeuLysSerLeuAla 286
DB 5 AATGAGGTTGAGAAATTCGGAATATCAACTACTTTCATCAATGTTAGTTGAGAGTTGGCT 64
OY 287 TyrPheMetValAlaProThrLeuCysTyGlnProSerTyTrpArgSerAlaCysIle 306
DB 65 TACTTCAGAGTTGCTCCCAACTTATGCTATACGCTTACCTATCCCTGCTGCATCATAT 124

```


REFERENCE
1 (bases 1 to 813)
AUTHORS
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
JOURNAL
Unpublished
COMMENT
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 16 Row: H Column: 13
Seq primer: gta aac cga cgg cca gtc.
Location/Qualifiers
1. 813

FEATURES
source
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNE02J19"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEB"
/note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2:
XhoI; 24 hrs after inoculation with Rice Blast (Che 86061
ORIGIN

BASE COUNT 211 a 147 c 165 g 290 t

Alignment Scores:
Pred. No.: 1,39e-82 Length: 813
Score: 959.50 Matches: 169
Percent Similarity: 76.95% Conservative: 38
Best Local Similarity: 62.83% Mismatches: 55
Query Match: 34.63% Indels: 7
DB: 14 Gaps: 1

US-09-623-514a-2 (1-520) x CB651490 (1-813)

OY 156 MetLysTyrGlyTrpLeuIleArgThrAspPheSerSerArgSerLeuArgAsp 175
DB 6 TTGCAGTATGCGCTTATTAATAGAGCTGGGTTTGGTTTATGATTAATCATTCGCGGAC 65
OY 176 TrpProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrVal 195
DB 66 TGGCCACTTCTATATGTTGTCTTAATGTCCTCCGCTTCCCGCTGGTGCATTTGCAGTT 125
OY 196 GluLysLeuValLeuGlnLysTyrTrpIleSerGluProValIlePheLeuHisIleIle 215
DB 126 GAAAGCTTGGCATTTAACAATGTTATCTGATGCTGCTACTGCTCCCATATCTTC 185
OY 216 IleThrMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPhe 235
DB 186 CTTCACACACCGCAAAATGATATACCAAGCTGCTGATTCCTTAAGTGTGATTCGCGATT 245
OY 236 LeuSerGlyValThrLeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTyr 255
DB 246 TTGTCGGCGCTTTTGTGATATTTATTTGCTGATTCGCTGCAAGCTGTCATCTTTT 305
OY 256 AlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGlu 275
DB 306 GCACATACCAACCATGATATTAAGCAACGATGCGCGCAAGAGGTTGATTAATGAA 365
OY 276 ValSerTyrTyr-----ValSerLeuLysSerLeuAlaTyrPhe 288
:::||||| ::||| ::||| ::|||

DB 366 CTAGCAGCTGTCAGATGATTAATTTACAACTCCCACTTAGGAATCTAATACTTC 425
OY 289 MetValAlaProThrIleuCysTyrGlnProSerTyrProArgSerAlaCysIleArgLys 308
DB 426 ATGATGCTCTTCACTCTTATTCACCAAGCTATCCCGAAGCTTATGTTAGAAA 485
OY 309 GlyTrpValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIle 328
DB 486 GGTGGCTGATGTCGAATATATCTGACTGATGTTACTGCTTCAAGGCTTCATT 545
OY 329 IleGlnIleThrIleAsnProIleValAlaArgAsnSerLysHisProLeuLysGlyAspLeu 348
DB 546 ATGAGCAATACATTAATTCCAATTTGTTGATTCAGATTCATCATGAAAGGAGGACTC 605
OY 349 LeuTrpAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrValTrpLeuCys 368
DB 606 CTAAATCTCTAGAGACTGTTTGAACCTCTCATTTACCAATGTTTACCTGCTGCTTGC 665
OY 369 MetPheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLeuCysPheGly 388
DB 666 ATGTTCTATGCTTTTCCATCTCTGTTAGTATGTTGCTGATTCGATTGCTT 725
OY 389 AspArgGluPheThrLysAspTrpTrpAsnAlaLysSerValGlyAspTyrTrpArgMet 408
DB 726 GACCGGTAATCTTACCAAAAGATGCTGGAATGCCAATAATGATGATGAGAGAAA 785
OY 409 TrpAsnMetProValHisLysTrpMet 417
DB 786 TGGATATGCTCTGATCAATAATGGTT 812

RESULT 7
CB660636 851 bp mRNA linear EST 09-APR-2003
LOCUS OSJNE02J19.f OSJNE02J19 5', mRNA sequence.
DEFINITION
Accession
CB660636.1 GI:29664361
VERSION
EST.
KEYWORDS
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Enarthroidae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 851)
Jantasuriyarat,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
JOURNAL
Unpublished
COMMENT
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aac cga cgg cca gtc
BACKWARD: gga aac agc tat gac cat g
Plate: 02 Row: J Column: 19
Seq primer: gta aac cga cgg cca gtc.
Location/Qualifiers
1. 851

FEATURES
source
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNE02J19"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_lib="OSJNEB"

```

OY 387 phegliaaparggluphetrrlylsasprtrpasaalalysserValglAspyrrtrp 406
    |||||||
DB 365 TTGGGGATCGGAGTTTATTAAGATTGGTGAACGCAAAAACATTCATGATGATTGG 424
OY 407 ArgmetrPasmaProValHisLysTrpMetValaGHisIleTyPheProCysLeu 426
    |||||||
DB 425 AGACTTTGGANATGCCCTGACATAGTGAGTGTCGTCACATCATATTCCACGCTTA 484
OY 427 ArgSerLysIleProLysThrLeuAlaIleIleIleAlaPheLeuValSerAlaValPhe 446
    |||||||
DB 485 AGAATGCAATCACTAAGAGAGTGCATGATCTCTTCTTATATCATCTCTGTTTC 544
OY 447 HIsGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuLysIle 466
    |||||||
DB 545 CATGGCTGTGTATTGGTCTGCTTCTGCGCTATTCAGTTTGGCATCTCTTGGAAAC 604
OY 467 MetPheGlnValProLeuValPheIleThrAsnTrpLeuGlnGluArgPhe---GlySer 485
    |||||||
DB 605 ATGTTTCAGATTCCTTGGTCTACTACAGAACTTCCTGCACAAACAGTTCAAAACCTCG 664
OY 486 ThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCysVal 505
    |||||||
DB 665 AATGGGCAACATGACATCTGCTGCTTCTTCTGCAATGTTGCAACCATG-TGTGTG 723
OY 506 LeuLeu 507
    ||||||
DB 724 CTCTGTG 729

```

```

RESULT 5
LOCUS CB660599 872 bp mRNA linear EST 09-APR-2003
DEFINITION OSJNEB02118.f OSJNEB Oryza sativa (japonica cultivar-group) cDNA
VERSION CB660599
KEYWORDS EST
SOURCE CB660599.1 GI:29664324

```

```

ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehharidiales; Oryzae; Oryza.
1 (bases 1 to 872)
Jantassurlyaret,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
Unpublished
Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ
85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aac cga cgg cca gtg
BACKWARD: gga aac agc tat gac cat g
Plate: 02 row: 1 column: 18
Seq primer: gta aac cga cgg cca gtg.

```

```

FEATURES
Source
Location/Qualifiers
1..872

```

```

1..872
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEB02118"
/tissue_type="Leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_id="OSJNEB"
/note="Vector: pBluescript II KS+. Site.1: EcoRI, Site.2:
XhoI; 24 hrs after inoculation with Rice Blast (C9240-1)"

```

```

BASE COUNT 227 a 174 c 172 g 299 t
ORIGIN

```

Alignment Scores:

```

Pred. No.: 3,2e-83 Length: 872
Score: 966.50 Matches: 162
Percent Similarity: 87.05% Conservative: 33
Best Local Similarity: 72.32% Mismatches: 28
Query Match: 34.88% Indels: 1
DB: 14 Gaps: 1

```

```

US-09-623-514a-2 (1-520) x CB660599 (1-872)

```

```

OY 293 ThrLeuCysTyrglnProSerTrpProArgSerAlaCysIleArgLysGlyTrpValAla 312
    |||||||
DB 6 ACACCTGTATACAGCAAGATATCCCGCACTCATGTTAGAAAAGTTGGCTGAT 65
OY 313 ArgGlnPheAlaLysLeuValIlePheTrpGlyPheMetGlyPheIleIleGlnTrp 332
    |||||||
DB 66 CGTCAAAATTAATCTGACTGATCTTACTGCTTCAAGGCTTCATTTAGCAATAC 125
OY 333 IleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeuLeuTrpAlaIle 352
    |||||||
DB 126 ATTAATCCATTTGTTGATTCACGATCCATTCATTAAGAGAGACTCTTAATGCTGTA 185
OY 353 GluArgValLeuLysLeuSerValProAsnLeuTrpValTrpLeuCysMetPheTrpCys 372
    |||||||
DB 186 GAGACTGTTTGAACCTCATACCAAAATGTTACCTGTGCTTGCATGTTCTATGCT 245
OY 373 PhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLysCysPheGlyAspArgGluPhe 392
    |||||||
DB 246 TTTTTCACATCTGTGTAGTAACTGCTGTGAGATTTCTGATTTGGTACCGTAATTC 305
OY 393 TyrLysAspTrpTrpPasmaAlaLysSerValGlyAspTrpArgMetTrpAsnMetPro 412
    |||||||
DB 306 TCAAAAGATTTGGTGAATCAAAACAAATGATGATTTGAGAAAATGGAATATGCTT 365
OY 413 ValHisLysTrpMetValArgHisIleTyPheProCysLeuArgSerLysIleProLys 432
    |||||||
DB 366 GTACATAAATGGGTGTGTCGCAATTTACTTCTTGCATGCGAATATGTTATCAAG 425
OY 433 ThrLeuAlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLeuCysIleAla 452
    |||||||
DB 426 GAAGTGTGCTGTGATGATCAATCTGTTTGTGCGGATCTCCAGATGATATGTCGCT 485
OY 453 ValProCysArgLeuPheLysLeuTrpAlaPheGlyIleMetPheGlnValProLeu 472
    |||||||
DB 486 GTTCCCTGGCGCATCTCAAGTTCGCGCATCTTAGGAATATGCTACAGATCCCTTT 545
OY 473 ValPheIleThrAsnTrpLeuGlnGluArgPheGlySerThr---ValGlyAsnMetIle 491
    |||||||
DB 546 ATCGTATTCAGACATGACTCAAAAGTAATTCAGAGATACATGATGGTGGCAACATGATA 605
OY 492 PheTrpPheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTrpTrpHisAsp 511
    |||||||
DB 606 TTTTGTCTTTTCTGATCTATGAGGCGAGCAATGTGCTTCTCTGATCATCATGAT 665
OY 512 LeuMetAsnArg 515
    |||||||
DB 666 GTGATGACACAG 677

```

```

RESULT 6
LOCUS CB651490 813 bp mRNA linear EST 08-APR-2003
DEFINITION OSJNEB16H13.f OSJNEB Oryza sativa (japonica cultivar-group) cDNA
VERSION CB651490
KEYWORDS EST
SOURCE CB651490.1 GI:29646483
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```

ORIGIN

Alignment Scores:

Pred. No.: 7,69e-93 Length: 1181
 Score: 1067.50 Matches: 192
 Percent Similarity: 79.71% Conservative: 28
 Best Local Similarity: 69.57% Mismatches: 47
 Query Match: 38.52% Indels: 9
 DB: 11 Gaps: 3

US-09-623-514a-2 (1-520) x AY110660 (1-1181)

DB 253 ValSerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAspLys--- 271
 12 GTCTCTATCATACAAATATGATATAGGATATGTCACAAATCTGAGAAAGGT 71
 DB 272 -----AlaAsnProGluValSerTyrTyrValSerLeuLysSerLeu 285
 72 GGTGCATAGGAATATATGTCATCTGAGATATGAAAGATCCAACTTTAAAGTCTA 131
 DB 286 AlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCys 305
 132 GTGTACTCATGTTGGGCCCAACACTTGTACCAGCACTATCTCTCAACATCATGT 191
 DB 306 IleArgLysGlyTyrValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMet 325
 192 ATTAGAAAGGTTGGTGGTACCCAGCAACATCATTAAGTGGTGGNNNNNACAGGCTTGATG 251
 DB 326 GlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLys 345
 252 GGGCTTCAATTAAGCAATATATTAACCAATGTTGAAGATTCACCAATCCACTGAAA 311
 DB 346 GlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrVal 365
 312 GCGAANNNNNGAATGTATAGAAAGGCTTTAAACTCTCAGTGGCCCACTTATATGTA 371
 DB 366 TrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeu 385
 372 TGGCTTGCATGTTCTATTCCTNNNNNNNCTTATGAGTGAACATTTGAGCTGCACTCTC 431
 DB 386 CysPheGlyAspArgGluPheTyrLysAspTyrTrpAsnAlaLysSerValGlyAspTyr 405
 432 TGTTCGTCAGCCGTGAATCTATTAAGACTGGGGAATCCCAAACTGTTGAAGACTAC 491
 DB 406 TrpArgMetTrpAsnMetProValHisLysTrpMetValArgHisLysLysPheProCys 425
 492 TGGAGGATGTGAACAATGCTGTTCAATGATCATCAGACACATATATTTCCATGT 551
 DB 426 LeuArgSerLysIleProLysThrLeuAlaIleIleIleAlaPheLeuValSerAlaVal 445
 552 ATTAGCAAGAGCTTTTCCAGGGGTGAGCTATCTATCTGCTTCTGTTTCAGCTGTA 611
 DB 446 PheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGly 465
 612 TTCCATAGATATGATATGCGGTGCGGCGACATTTTCAAAATCTGCGCATTTTCTGGG 671
 DB 466 IleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySer 485
 672 ATCATGTTTCAGATACGCTGTTGATCTTGAACAAGATATCTCCATGCTAGCTTCAACAT 731
 DB 486 Thr---ValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnPrometCys 504
 732 GTAATGTTGGCAACATGATATTTTGGTTC---TTCAGTATAGTCGAGACAGCCGATGTGT 788
 DB 505 ValLeuLeuTyrTyrHisAspLeuMetAsnArgLysLysSerMetSer 520
 789 GTCTTCTATATACATGAGAGCTATGATGACAGGAGCCAGGCAAT 836
 DB 789 GTCTTCTATATACATGAGAGCTATGATGACAGGAGCCAGGCAAT 836
 RESULT 4
 BI422326 729 bp mRNA linear EST 16-AUG-2001
 LOCUS EST552392 tomato callus, TANK Lycopersicon esculentum cDNA clone
 DEFINITION GLEC69N10 5' end, mRNA sequence.

ACCESSION BI422326
 VERSION BI422326.1 GI:15196624
 KEYWORDS EST
 SOURCE Lycopersicon esculentum (tomato)
 ORGANISM Lycopersicon esculentum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 asterids; lamiales; Solanales; Solanaceae; Solanum; Lycopersicon.
 REFERENCE 1 (bases 1 to 729)
 AUTHORS Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
 Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Roming,
 C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.J.
 Generation of ESTs from tomato callus tissue
 JOURNAL Unpublished
 COMMENT Contact: CUGI
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Email: <http://www.genome.clemson.edu/orders/index.html>
 Location/Qualifiers
 1..729
 source
 FEATURES
 source
 /organism="Lycopersicon esculentum"
 /mol_type="mRNA"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /clone="GLEC69N10"
 /tissue_type="callus"
 /dev_stage="25-40 days old"
 /lab_host="Xil-Blue MRP"
 /clone_id="tomato callus, TANK"
 /note="Vector: pBluescript SK(-); Site.1: EcoRI; Site.2:
 XhoI; Supplier: GibcoBRL laboratory; GLEC - Cotyledons
 of seedlings 7-10 days post-germination were excised, cut
 at both ends and placed on MS medium with no selection.
 Mixed callus was harvested at 25 and 40 days and included
 undifferentiated masses. Tomato Callus EST library"

BASE COUNT 177 a 142 c 160 g 250 t
 ORIGIN

Alignment Scores:

Pred. NO.: 2.5e-88 Length: 729
 Score: 1018.00 Matches: 185
 Percent Similarity: 85.95% Conservative: 23
 Best Local Similarity: 76.45% Mismatches: 33
 Query Match: 36.74% Indels: 2
 DB: 12 Gaps: 1

US-09-623-514a-2 (1-520) x BI422326 (1-729)

DB 267 AsnAlaIleAspLysAlaAsnProGluValSerTyrTyrValSerLeuLysSerLeuAla 286
 5 AATGAGGCTGAGATATCCGAATCACTACCTTCAATGTTATTTTCAGAGTTGCT 64
 DB 287 TyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIle 306
 65 TACTTCATGTTGTTCCAACTTATGCTATACGTTGATCTCTCCTCTGCTCATCAT 124
 DB 307 ArgLysGlyTyrValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGly 326
 125 CGAAGGTTGGTGGCGCCGCAACTCATCAAGCTGTAATTTTACAGGATTAATGGA 184
 DB 327 PheIleIleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGly 346
 185 TTATCATTTGAGCGATATATTAACCGATTTGGCAAGCTCCACACATCCATTGAAGA 244
 DB 347 AspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrValTrp 366
 245 AACCTTTATACGCAATGAGAGGCTATTTGAAGCTTTCAGTTCCAAATTTATATGTCGG 304
 DB 367 LeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLysCys 386
 305 CTCTGCATGTTCTACAGCTCTTTCATCTTGGTTAAATATCTTGGCAAGGTTCTGCA 364

TITLE Generation of a set of potato cDNA clones for microarray analyses
JOURNAL Unpublished
COMMENT On Jun 10, 2002 this sequence version replaced gi:21365236.
REFERENCE 1 (bases 1 to 961)
AUTHORS Bulli,C.R., Hatt,A., Baker,B., Tanksley,S., Fry,W., Smart,C., Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and Karamycheva,S.A.

ORIGIN	324 a	205 c	181 g	251 t
BASE COUNT				
Alignment Scores:				
Pred. No.:	1.85e-101	Length:	961	
Score:	1155.00	Matches:	209	
Percent Similarity:	86.72%	Conservative:	26	
Best local Similarity:	77.12%	Mismatches:	32	
Query Match:	41.68%	Indels:	4	
DB:	13	Gaps:	3	
US-09-623-514A-2 (1-520) x BQ510367 (1-961)				
QY	254	SerTyrAlaHisThrSerTyrAspIleArgSerIleuAlaAsnAlaAspLysAla---	272	
Db	961	TCTATATGCACATACAAATTATGATATGAGACGCTTGCAAGCTGTGTGATAGAGGTGA	902	
QY	273	AsnProGluVal-----SerTyrTyrValSerIleuLysSerIleuAlaTyrPheMetAl	290	
Db	901	AAATCCGAAACCACTACTCTTAACAAATGATAGTTTCAAGAGCTTTGGCTTCAATGCTT	842	
QY	291	AlaProThrLeuGlyTyrGlnProSerTyrProArgSerAlaCysIleArgLysGlyTyr	310	
Db	841	GCTCAACTTATATGCTATACAGCTTAAGCTATCTCGCTCGATCATTCGGAAGGTTGG	782	
QY	311	ValAlaArgGlnPheAlaLysIleuValIlePheThrGlyPheMetGlyPheIleIleGlu	330	
Db	781	CTGGCCGCCCAACATCAGCAAGCTGATATTTCACAGATTAATGGGATTATCATTTGAG	722	
QY	331	GlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeuLeuTyr	350	
Db	721	CAGATATTTAAACCCGATTTGCCGAGATCCACACATTCATTGAGGAACCTTTTAAAC	662	
QY	351	AlaIleGluArgValIleuLysLysLeuSerValProAsnLeuTyrValIleTyrPheCysMetPhe	370	
Db	661	GCCATCGAAGAGGATTTGAGCTTTCAGCTTCATTTTATATATGCTGGCTGTGATGTTTC	602	
QY	371	TyrCysPhePheHisLeuTyrPheAsnIleuAlaGluLeuLeuCysPheGlyAspArg	390	
Db	601	TACAGCCTCTTTCATCTTTGGCTAAATATACCTTCGGAATATTCGTGCATTTGGGAGTCGT	542	

FEATURES	SOURCE
391 GlubeheryrlvsaspttpttprasnalsylserValaGlyaspyrrtpragmetprasn	410
541 GAGTTCACAAAGAGTGGTGGAAATGCAAAAAACAATGATGATATGAGACTTTGGAT	482
411 MetproValaHlsVsttptmetValaArgHlsIleTyrPheproCysIeaArgserLysIle	430
481 ATGGCTGTGACATAAGTGGATGGTTCGGCACATTTATTTTCATCTTAAAGGAATGGCATA	422
431 ProLysThrIleuAlaIleIleIleAlaPheLeuValserAlaValaPheHISGluLeuCys	450
421 CCTAAGGAGAGTTCGAAGGTCGATCTCTTCTTTATCTGCTGTTTCATGACACTTGAAT	362
451 IleAlaValaProCysArgIleuPheLysLeuTrrpIaPheLysGlyIleMetPheGluVal	470
361 ATTCCTGTCTCTTGTGGCTATTCAGCTTTGGGACATCTTGGAAATCATGTTTCACAATT	302
471 ProLeuValaPheIleThrAsnTyrLeuGlnGluArgPhe--GlySerThrValGlyAsn	489
301 CCCCTGGTGCATACCCAGCAACTTCTCTGCAAAACAAGTTCAAAAACCTGATGTGGCAAC	242
490 MetIlePheTrrpPheIlePheCysIlePheGlyGlnProMetCysValaLeuTyrTyr	509
241 ATGACATCTTCGTGTTTCTTTCTGCATGTTGGTTCACCAACCAATGTGTGCTTCTGTATTAAC	182
510 HisAspleuMetAsnArgLysGlySerMetSer	520
181 CATGATGTGATGAATAGAAATGATGTTCAAGT	149
RESULT 3	
AY110660	
LOCUS	
DEFINITION	Zea mays CL739_1 mRNA sequence.
ACCESSION	AY110660
VERSION	AY110660.1 GI:21215250
KEYWORDS	HTC.
SOURCE	
ORGANISM	Zea mays
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE	1 (bases 1 to 1181)
AUTHORS	Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE	Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes
JOURNAL	Unpublished (2002)
REFERENCE	2 (bases 1 to 1181)
AUTHORS	Coe,E.H.
TITLE	Direct Submission
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
REFERENCE	if you are interested in getting corresponding physical clones, these are publicly available from zmdb and may be found by BLAST searching at NSL, www.maizemap.org ; ZMDb, www.zmdb.iastate.edu ; RGR, www.rgr.org ; or NCBI, www.ncbi.nlm.nih.gov . When the source of the maize cDNA sequences is either Virginia Walbro, Stanford or Pat Schnable, Iowa State, then clones may be requested from ZMDb: www.zmdb.iastate.edu .
COMMENT	location/Qualifiers
	1..1181
	/organism="Zea mays"
	/mol_type="mRNA"
	/db_xref="MaizeDB:632842"
	/db_xref="taxon:4577"
	/clone_lib="Maize Mapping Project/Dupont Cornusens library"
	/note="this sequence is part of a project of EST assemblies resulting from the application of public configs to seed Dupont configs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
BASE COUNT	322 a 240 c 275 g 327 t 17 others

BASE COUNT	322 a	240 c	275 g	327 t	17 others
------------	-------	-------	-------	-------	-----------

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 31, 2003, 05:00:54 ; Search time 2515 Seconds
(without alignments)
5025.180 Million cell updates/sec

Title: US-09-623-514a-2
Perfect score: 1
Sequence: 1 MAILDSAGVTTYTENGSGGF.....QPMCVLYHYHDMNRKSGMS 520

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 22781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+pn.model -DEV-xlp
-Q/csn2.1/USPTO-SPPOOL/US0963511/runat.29082003.152200.62/app.query.fasta.1.711
-DB-EST -OFMT-fastap -SUFFIX-trst -MINMATCH=0.1 -LOOPTOL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX-biosum62 -TRANS-human40.cdi -LIST=45
-OCALIGN=200 -THR-SCORE-pct -THR-MAX=100 -THR-MIN=0 -ALIGN=15 -MODE-LOCAL
-USER=US0963511-ECGN.1.1.4237-@runat.29082003.152200.62 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEUDERY -NEG-SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV-TIMEOUT=120 -WARN-TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST:
1: em_estba:*
2: em_estlum:*
3: em_estlin:*
4: em_estnu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrl:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_prod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1.*

29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1279	46.2	1572	11	AY105372
2	1155	41.7	961	13	B0510367
3	1067.5	38.5	1181	11	AY110660
4	1018	36.7	729	12	BI422326
5	966.5	34.9	872	14	CB650599
6	959.5	34.6	813	14	CB651490
7	943.5	34.0	851	14	CB660636
8	902	32.6	586	13	B0148998
9	835.5	30.2	741	13	BQ998922
10	832.5	30.0	729	12	BJ322102
11	765	27.6	539	12	BI422212
12	762.5	27.5	606	12	B316561
13	734.5	26.5	855	9	AM349274
14	731.5	26.4	656	14	CB619077
15	727.5	26.3	540	9	AJ470192
16	726	26.2	572	13	B0124305
17	725.5	26.2	820	14	CB619078
18	689	24.9	561	9	AM035727
19	687.5	24.8	862	10	B6321213
20	683.5	24.7	626	9	AV926912
21	653	23.6	816	29	B2503532
22	652.5	23.5	629	9	AA042298
23	635.5	22.9	739	14	CB646515
24	635.5	22.9	794	14	CB651491
25	625.5	22.6	655	13	B0998311
26	600	21.7	673	13	BQ998554
27	598.5	21.6	480	9	AJ470191
28	595.5	21.5	480	9	AJ470190
29	589	21.3	770	14	CB646514
30	588	21.2	746	28	BH466416
31	583.5	21.1	736	14	CB620474
32	579.5	20.9	828	13	B0861203
33	575.5	20.8	865	14	CB660600
34	571	20.6	1064	12	BM476315
35	562.5	20.3	555	12	BJ268713
36	559	20.2	1043	12	BM919422
37	551	19.9	656	13	B0862755
38	546.5	19.7	515	9	AV925760
39	542.5	19.6	412	14	CB663593
40	534.5	19.3	807	14	CB660637
41	533	19.2	627	14	CA784021
42	532.5	19.2	748	14	CB620472
43	531.5	19.2	508	13	B0009402
44	529.5	19.1	959	9	AL520752
45	528.5	19.1	968	13	B0670343

ALIGNMENTS

RESULT 1
AY105372
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE

AY105372
Zea mays
PC0148220 mRNA sequence.
AY105372
AY105372.1 GI:21208450
HTC.
Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
Clade; Panicoideae; Andropogoneae; Zea.
1572 bp mRNA linear HTC 16-OCT-2002
1 (bases 1 to 1572)

```

0y      498 epehglycIn 501
      ::: |||
Db      3015 TTAATTCAA 3024

Search completed: August 31, 2003, 07:17:10
Job time : 132 secs

```

Db 2238 TTCAGCAGATCGTTTGTAAATGAAAGCCCACTATTGTGAGAGACAGACTGCCCTGGG 2297
 268 -----AlaAlaAspLysAlaAsn-----ProGluValSerTyrTyrVa 280
 2298 TACTAATTCAGCTAGAGAAATCAAGACAGCTTCCATACCTACAGTCAACAGATATT 2357
 280 1SerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuGlyGlnProSerTyr 300
 2358 G-----TACTCTTATTTGCTCTACCTCTACCTCTATCTACCTGACAGCTA 2399
 300 PProArgSerAlaCysIleArgLysGlyTyrIleValAlaArgGlnIleAlaLysLeuValI 320
 2400 TCCACAGGATCCCACTGTAAATAGGGGTTATGTGTATGAAGTTAGCAGAGCTTTGG 2459
 320 ePheThrGlyPheMetGlyPheIleLeuGlnTyrIleAsnProIleValAlaArgAsnSe 340
 2460 TTGCTTTTTCATGTGTACATCATCTTGAAGAGCTTGCCCTGTTGTTGGAAAT 2519
 340 rLys---HisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSe 359
 2520 CAAACAGAGCCCTTCAGCCCT-----CGTGTTCGTGCTCATG 2558
 359 rVal-----ProAsnLeuTyrValIleProLeuGlyMetPheTyrCysPhePh 374
 2559 TGTATTTAACTCCATCTTCCAGAGTGCTGATTCCTCTCTCTACTTTTTCCTTTT 2618
 374 eHisLeuTyrPheAsnIleLeuAlaGluLeuLeuGlyAspArgLysPheTyrLys 394
 2619 GCAGTGGCTGCTCAATGCTTGTGCTGATGTACGCTTGTGTACAGAGATGTTCTA 2678
 394 sAspTyrTyrPheAsnAlaLysSerValGlyAspTyrTyrArgMetTyrPheAsnMetProValI 414
 2679 GGATTTGGTGAACCTCCAGCTACTCTCACTATTAAGAACCTGGAATGTGGTGCTCA 2738
 414 sLysTyrMetValArgHisIleTyr-----PheProCysLeuArgSerLysIleProLys 432
 2739 TGACTGGTATATTAAGTATGCTTACAGAGACTTCTCTGTTTCTTCCAGAGATTC 2798
 432 sThrLeuAlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLeuCysIleAl 452
 2799 ATCGTGCCATGTAGTGTCTTGTGCTATCTGCTGTATACAGAGATATGCTTGGC 2858
 452 aValProCysArg-----LeuPheLysLeuTyrAlaPheLeuGlyI 466
 2859 TGT---TGCCTTGAGCTTTTCTATCCGCTGCTGTCTGCTCATGATCTTGGAA 2915
 466 eMetPheGlnValProLeuValPheIleThrAsnTyrLeuGlnIleArgPheGlySerTh 486
 2916 GCGTTTCAC-----TTCATTTCAATGATATGCGAATAAG-----CC 2954
 486 rValGlyAsnMetIlePheTyr-----PheIlePheCysAl 498
 2955 GATTGGATGTCTGATGTGAGACTTCTTCTTGGGCAATGAGAGCTTACTCTGCTT 3014
 498 ePheGlyGln 501
 3015 TTATTCCTCAA 3024

RESULT 15

US-08-509-187D-2
 : Sequence 2, Application US/08509187D
 : Patent No. 5834283
 : GENERAL INFORMATION:
 : APPLICANT: Chang, Ta-Yuan and Chang, Catherine C.Y.
 : TITLE OF INVENTION: ACYL Coenzyme A: Cholesterol Acyltransferase
 : NUMBER OF SEQUENCES: 9
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: LAHIVE & COCKFIELD, LLP
 : STREET: 28 State Street
 : CITY: Boston
 : STATE: Massachusetts
 : COUNTRY: USA
 : ZIP: 02109

; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/509,187D
 ; FILING DATE: 31-JUL-1995
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lamport Hamillite, Ann
 ; REGISTRATION NUMBER: 34,858
 ; REFERENCE/DOCKET NUMBER: DCI-033cpdv
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (617)227-7400
 ; TELEFAX: (617)742-4214
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 4079 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; US-08-509-187D-2

Alignment Scores:

Pred. No.: 5,57e-34 Length: 4079
 Score: 405.50 Matches: 148
 Percent Similarity: 41.36% Conservative: 77
 Best Local Similarity: 27.21% Mismatches: 167
 Query Match: 14.63% Indels: 153
 DB: 2 Gaps: 24

US-09-623-514a-2 (1-520) x US-08-509-187D-2 (1-4079)

QY 17 GlyGlyGluPheValAspLeuAspArgLeuArgLysSerArgSerAspSerSer 36
 1675 GGCAGTCACTTGTATGATTTGTGACCAATCTCATTTGAAAGTACGATCTTGTGATTA 1734
 QY 37 AsnGlyLeu-----LeuLeuSerGlySerAspAsnAsnSerProSer 50
 1735 GGTGGTGCGCTCCACAACTTTCTGTCTTGAGAGAGAAACAAACATAGAGCGC 1794
 QY 51 AspAspValGlyAlaProAlaAspValArgAspArgIleAspSerValAlaAsnAspAsp 70
 1795 AAGATTTGAGAGACCTCCAGAA----- 1818
 QY 71 AlaGlnGlyThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyAspAsnAsnGlyGly 90
 1819 ---CAAGAAAGATTTTATT----- 1836
 QY 91 GlyArgGlyGlyGlyGluGlyArgGlyAsnAlaAspAlaThrPheThrTyrArgProSer 110
 1836 ----- 1836
 QY 111 ValProAlaHisArgAlaArgLysSerProLeuSerSerAspAlaIlePheLysGln 130
 1837 -----GCAAGCGCGCTCTCTCTTA-----GATGAACCTGCTTGAAGTG 1872
 QY 131 SerHis---AlaGlyLeuPheAsnLeuGlyValValLeuIleAla---ValAsnSer 148
 1873 GACCACTACGACAAATATATACATGTTTATTCCTCCCTCATCTTATCTGAGC 1932
 QY 149 ArgLeuIleLeuGlu-AsnLeuMetLysTyrGlyTyrPheLeuArgThrAspPheTyrPh 168
 1933 ACATTTGTAGTAGATATCATGATGAGGAAGGCTGTGCTGCAAGTACG----- 1984
 QY 168 eSerSerArgSerLeuArgAspTyrProLeuPheMetCysLysIleSerLeuSerIlePh 188
 1985 -----CCCTCTGTCTTATGCAAT-----TTGGCAAA--TT 2012

```

QY 329 IlegluGluTyrIleAsnProIleValArgAsnSerLys---HisProLeuLysGlyAsp 347
      |||::: |||::: |||::: |||
Db 2417 TTTGAAGGCTTTGGCCCTTGTTCGGAATATCAACAGGACCTTCACGGCT--- 2473
QY 348 LeuLeuTyrAlaIleGluArgValLeuLysLeuSerVal-----ProAsn 362
      ||| ||| ||| ||| |||
Db 2474 -----CGTGTCTGGTCTGATGATGATTTTACCTTCATCTTGGCCAGGT 2515
QY 363 LeuTyrValTyrLeuLysMetPheTyrCysPhePheHisLeuTyrLeuAsnIleLeuAla 382
      ::::: |||::: ||| ||| ||| ||| |||
Db 2516 GTGCTGATTCCTCTCTTACTTCTTTTGGCTTTTTCACAGTGGCTGATGCTTTGGCT 2575
QY 383 GluLeuLeuLysPheLeuLysArgGluPheTyrLysAspTyrTrpAsnAlaLysSerVal 402
      |||::: ||| ||| ||| ||| ||| ||| ||| |||
Db 2576 GAGATGTACGGTCTTGGTACAGAGATTTCTATAGAGATTTGGGAGACTCCACGTCATAC 2635
QY 403 GlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHisLysTyr 422
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 2636 TCCCACTATATATAGAACCTGGATGTGGTGCTCCATGACTGGCTATATATCTATGCTTAC 2695
QY 423 -----PheProCysLeuArgSerLysIleProLysTrpLeuAlaIleIleAlaPhe 440
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2696 AAGGACTTCTCTGGTCTTCTCCAGAGATTCAAATCTGCTCCATGTAGCTGTCTT 2755
QY 441 LeuValSerAlaValPheHisGluLeuLysIleAlaValProCysArg----- 456
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2756 GCTGATACGCTGTAGTACACGAAATATGCTTGGCTGT---TGCCTGAGCTTTTCTAT 2812
QY 457 -----LeuPheLysLeuTyrPheAlaPheLeuGlyIleMetPheGluValProLeuValPhe 474
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2813 CCCGCTCTTCTGCTCTTCTCATGTTCTTGGATAGCTTTCAC-----TTC 2860
QY 475 IlePheAsnTyrLeuGluGluArgPheGlySerThrValGlyAsnMetIlePheTyr--- 493
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2861 ATTGTCAAGATAGATGGAAGAAAG-----CCGATTTGGAAATGTTCTGATGTGAGCT 2911
QY 494 -----PheIlePheCysIlePheGlyGln 501
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2912 TCTCTTTTCTGGGCAAGGAGCTTACCTGCTTTTATTTCA 2956

RESULT 14
US-08-121-057-2
: Sequence 2, Application us/08121057
: Patent No. 5484727
: GENERAL INFORMATION:
: APPLICANT: CHANG, TA-YUAN
: APPLICANT: CHANG, CATHERINE C. Y.
: TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
: TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LAHIVE & COCKFIELD
: STREET: 60 STATE STREET, SUITE 510
: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: ASCII Text
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/121,057
: FILING DATE:
: CLASSIFICATION: 800
: ATTORNEY/AGENT INFORMATION:
: NAME: LAMPORT HAMMITTE, ANN.
: REGISTRATION NUMBER: 34,858
: REFERENCE/DOCKET NUMBER: DCI-033CP
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 227-2700
: TELEFAX: (617) 227-5941

```

```

: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4079 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
US-08-121-057-2

Alignment Scores:
Pred. No.: 5 57e-34 Length: 4079
Score: 405.50 Matches: 148
Percent Similarity: 41.368 Conservative: 77
Best Local Similarity: 27.216 Mismatches: 167
Query Match: 14.638 Indels: 153
DB: 1 Gaps: 24

US-09-623-514a-2 (1-520) x US-08-121-057-2 (1-4079)
QY 17 GlyGlyGluPheValAspLeuAspArgLeuArgArgLysSerArgSerAspSer 36
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1675 GGCAGTCACCTTGATGATTTTGGACCAATCTGATTGAAAGACGACGATCATTAATAT 1734
QY 37 AsnGlyLeu-----LeuLeuSerGlySerAspAsnAsnSerProSer 50
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1735 GGTGGTGCGCTCTCACAACTTTCTGTTCTTGAAGGAGAGAAACCATAGACGG 1794
QY 51 AspAspValGlyAlaProAlaAspValArgAspArgIleAspSerValAlaAsnAspAsp 70
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1795 AAGGATTTAGAGACCTCCAGAA----- 1818
QY 71 AlaGluGlyThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyLysAspAsnGlyGly 90
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1819 ---CAAGGAAGACATTTTATTT----- 1836
QY 91 GlyArgGlyGlyGlyGluArgGlyAsnAlaAspAlaThrPheThrTyrArgProSer 110
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1836 ----- 1836
QY 111 ValProAlaHisArgArgAlaArgGluSerProLeuSerSerAspAlaIlePheLysGln 130
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1837 -----GCAGGCGCTCTCTTTA-----GATACCTGCTTGAAGG 1872
QY 131 SerHis---AlaGlyLeuPheAsnLeuLysValValIleLeuAla---ValAsnSer 148
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1873 GACCACTGACGAACATATATACATGTTATGCCCCCTCATCTCTTATCTTCACGC 1932
QY 149 ArgLeuIleIleGluAsnLeuMetLysTyrGlyTrpLeuIleArgThrAspPheTyrPhe 168
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1933 ACACTTGATAGATTACATGTGATGAAAGAGGCTGCTTCGCAAGTTACG----- 1984
QY 168 eSerSerArgSerLeuArgAspTrpProLeuPheMetCysCysIleSerLeuSerIlePhe 188
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1985 -----CCTCTGCTTATGCAAT-----TTGGCAAA-TT 2012
QY 188 eProLeuAlaAlaPheThrValGluLysLeuValLeuGlnLysTyrIleSerGluProVal 208
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2013 TCTTACCGTTGTTTGACCTGGTGATGATGTTCCGCTGTACATTT---TCAAGTCCCA 2069
QY 208 ValIlePheLeuHisIleIleIleThrMetThrGluValLeuTyrProValTyrValHis 228
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2070 TTTTCTGTTCAACATGGCGCACTGGCTATGCAAGAGTTCTCATCCCTGATCCGCTC 2129
QY 228 rLeuArgCysAspSerAlaPheLeu-----SerGlyValThrLeuMetLeuLeuThr 245
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2130 TCTCTTCCATGGCTTTCTTTCATGATCTTCCAGATGTGAGTTCAGTTTGGACCAAC 2189
QY 245 rCysIleValThrPheLysLeuValSerTyrAlaHisThrSer-----Tyr 260
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2190 ATATGTTGTC-----TTAGCATATCTGACACACAGCTCCGCTTCATCATTA 2237
QY 260 rAspIleArgSerLeuAlaAsn----- 267
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


APPLICANT: Chang, Ta-Yuan and Chang, Catherine C.Y.
 TITLE OF INVENTION: ACYL Coenzyme A: Cholesterol Acyltransferase
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LAHIVE & COCKFIELD, LLP
 STREET: 28 State Street
 CITY: Boston
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/509,187D
 FILING DATE: 31-JUL-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Lamport Hammitte, Ann
 REGISTRATION NUMBER: 34,858
 REFERENCE/DOCKET NUMBER: DCI-033cpdv
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)742-4214
 TELEFAX: (617)742-7400
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4011 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1397..3046
 US-08-509-187D-3
 Alignment Scores:
 Pred. No.: 9,16e-35 Length: 4011
 Score: 412.50 Matches: 148
 Percent Similarity: 42.46% Conservative: 80
 Best Local Similarity: 27.56% Mismatches: 173
 Query Match: 14.89% Indels: 136
 Gaps: 23
 US-09-623-514a-2 (1-520) x US-08-509-187D-3 (1-4011)
 QY 17 GYGLGIYUHPHEVALASPLEUASPARGLAUARGATRGATGYSERARGSERASPSER 36
 DB 1604 GGCAGTCACTTGAATGATTTGTGACCAATCTCATTTGAAGTCCGATCATTAAT 1663
 QY 37 ASNGIYLEU-----LEUUSERSGLYSERASPSANASNSERPROSER 50
 DB 1664 GTGGGTGCGCTCACAACCTTTCTGTCTTGAAGAGAGAAAACACCATAGAGCG 1723
 QY 51 ASAPAPVALGIALAASPLEUASVALARGASPARGLIEAPSERVALVALASNASPAP 70
 DB 1724 AAGGATTTAGAGCCCTCCAGAA-----LEU 1747
 QY 71 ALAAGNGIYTHRALAASPLEUASGLYASPSANASNGIYGLYASPSANASNGIYGLY 90
 DB 1748 ---CAAGAAAGATTTTATT----- 1765
 QY 91 GLYARGGLYGLYGLUGLYARGGLYASNAALASPARALATHRPHEATHYRARGPROSER 110
 DB 1765 ----- 1765
 QY 111 VALPROALAHASARGALARGGLUSERPROLEUSERSERASPALAILEPHELYSGIN 130
 DB ----- 1765

DB 1766 -----GCAAGCGCTCTCTCTTA-----GATGAACCTGTTGAAGTG 1801
 QY 131 SERHIS---ALAGIYLEUPHEASPLEUCYSVALVALLEULEALA---VALANSER 148
 DB 1802 GACCACATCAGACAAATATATACATGTTATATGCCCTCCATCTCTTATCCCTCACC 1861
 QY 149 ARGLEULEILEGLU-ASPLEUMETLYSTYGLYTRPLEUILEARGTHRAPSPHERP 168
 DB 1862 ACACCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1907
 QY 168 EASERASERASERLEUARG-ASPTTRPROLEUPHEMETSYSIIESEULEUSER-ILE 187
 DB 1908 -TCAGCTTCCTGCTTATGCTTTTATGCTTTTATGCTTTTATGCTTTTATGCTTTTAT 1966
 QY 188 PHEPROLEUALAALAPHERTHRYVALGLYSEUVALLEUGNLYSTYR----- 203
 DB 1967 ATGTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2026
 QY 204 -----IIESEGLUPROVALILE-----PHELEHISILEILEILE 216
 DB 2027 AGCAAGATGCTCATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2086
 QY 217 THIMETHRGLUVALLEUTYR---PROVALITYVALTHRLEUARGCYASPSERLA 235
 DB 2087 CAGATTGAGGTTCTAGGTTTGGACCAATATGTTGTTA-----GCATAT 2134
 QY 236 LEUSERGLYVALTHRLEUMETLEUTHRCYSILEVALTRPLEUYSLEUVALSER 255
 DB 2135 ACACGCGCACAGCTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2194
 QY 256 ALAHISLTHR-----SETRYASPLEARGSERLEUALAASNAALASPLYSALA 272
 DB 2195 GCCACATCATTTGTCAGAGAGAACGTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 2254
 QY 273 ASN-----PROGLUVALSERTRYTRYVALSERLEUYSERLEUALATYR 288
 DB 2255 AGCACTGTTCCAAATACCTACAGTCAACCAAGTATTTG-----TACTTC 2296
 QY 289 METVALALAPROTHRLEUCYSTYRGLNPROSETRYPROARGSERALACYSILEARG 308
 DB 2297 TTATTGCTCCATACCTTATATCATGCTGACAGCTATCCAGAAATCCACATGATGAG 2356
 QY 309 GLYTRPVALAALARGGLNPHALASLEUVALILEPHERGILYPHMETELYPHEILE 328
 DB 2357 GGTATGTCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2416
 QY 329 ILEGLUGINTYRILEASPROILEVALARGANSERLYS---HISPROLEUYSLE 347
 DB 2417 TTGAAAGGCTTTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2473
 QY 348 LEULEUTYRALAILEGLUARGVALLEUYSLEUSERVAL-----PROASN 362
 DB 2474 -----CGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2515
 QY 363 LEUTYRVALITRPLEUCYSMETPHETRYCSPHEPHEHISLEUTRPLEUASNILE 382
 DB 2516 GTGCTGATGCTCTCTTACTTCTTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2575
 QY 383 GLULEULEUCYSPHEGLYASPARGLUBHERYLYSASPTPTPTPASHALALYSER 402
 DB 2576 GAGATGTTACGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2635
 QY 403 GLYASPT 422
 DB 2636 TCCAACTTATATGACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2695
 QY 423 -----PHEPROCYSEUARGSERLYSILEPROLYSTHRLAALALEILELAPHE 440
 DB 2696 AAGACCTTCTCTCTGTTTTTCCCAAGAGATCAATCTGCTGCTGCTGCTGCTGCTGCT 2755
 QY 441 LEUVALSERALVALPHEHISGLULEUCYSILEALVALPROCYARG----- 456
 DB 2756 GCTGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2812

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/121,057
 FILING DATE:
 CLASSIFICATION: 800
 ATTORNEY/AGENT INFORMATION:
 NAME: LAMPORT HAMMITTE, ANN.
 REGISTRATION NUMBER: 34,858
 REFERENCE/DOCKET NUMBER: DCI-033CP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 227-2700
 TELEFAX: (617) 227-5941
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4011 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1397..3046
 US-08-121-057-3

Alignment Scores:
 Pred. No.: 9,16e-35 Length: 4011
 Score: 412.50 Matches: 148
 Percent Similarity: 42.46% Conservative: 80
 Best Local Similarity: 27.56% Mismatches: 173
 Query Match: 14.89% Indels: 136
 DB: Gaps: 23

US-09-623-514a-2 (1-520) x US-08-121-057-3 (1-4011)

```

OY 17 GYGLYGLUPLPHEVALASPENLEUAPRGLAARGARGVLSERARGSERASP 36
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1604 GGCAGCTGACTTTCATGATGATTTTGTGACCAATCTTCATGAAAGTCACATCATTTAGTATAT 1663
OY 37 ASNGLYLEU-----LEULEUSERGLYSERASPANSERPROSER 50
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1664 GGTGGGTGGCTGCTGCACACACTTTTCTGCTTGTGAAGAGAGAAAAACACATACAGG 1723
OY 51 ASPAPVALGVALAPROVALASPYVALARGSPATGILEASPSERVALVALASNASPAP 70
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1724 AAGGATTTGAGACACCTCCAGAA----- 1747
OY 71 ALANGNGLYTHRALASPLEUALLAGLYASPANASNGLYGLYASPANASNGLYGLY 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1748 ---CAAGCAACATTTTATTT----- 1765
OY 91 GLYATGGLYGLYGLUGLYGLUGLYASNALASPALATHRPHERTHYRARGPROSER 110
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1765 ----- 1765
OY 111 VALPROLAHLSARGARGVALARGGLUSERPROLEUSERSERASPALLLEPHELYSGLN 130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1766 -----GCAAGGGCTCTCTCTTA-----GATGACACTGCTTGAACG 1801
OY 131 SERHLS---ALAGLYLEUPHEASLEUCYSVALVALLEUALLLELA---VALANSER 148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1802 GACCAATCAGACAAATATATACATGTTTATGCCCTCCATCTCTTTATCCACAC 1861
OY 149 ARGLEULEILEGLU-ASNLEUMELYSTYGLYTRPLEUULEARGYTHRASPHERTRPH 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1862 ACACCTGTGATGATATCATGATGAGAGAGGCTGCTGAGT----- 1907
OY 168 ESERSEARGSERLEUARG-ASPTTRPROLEUPHEMELCYSLLESERLEUSER-ILE 187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1908 -TCACCCCTCTCTCTATGCTTTGCAAAATTTCTACCCGCTTGGACCTGCGAGAC 1966
OY 188 PHEPROLEUALALAPHERTHVALGULYSLEUVALLEUGLLYSTYR----- 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1967 ATGTTCTCTCTACATTTTCAGTTCCCTATTTTCTGTTTCAACATGTCGCGACCTGCTAT 2026
  
```

```

OY 204 -----LESERGLUPROVALLE-----PHEUHSILEILE 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2027 AGCAAGATTCATCCGCTGATCCGCTCTCTCTCTCCATGCGCTTCTTTATGATCTTC 2086
OY 217 THRMTHTHGLUVALLEUTYR---PROVALTYRVALTHLEUARGCYSASPERALAPHE 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2087 CAGATTGAGATTCTAGTTTGTGACCAACATATGTTGTCTTA-----GCATAT 2134
OY 236 LEUSERGLYVALTHLEUMELLEUETHYRCSILEVALTRPLEUYSLEUVALSERTYR 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2135 ACACGCCACCAAGCTCCCGGCTCATATATTCGACAGATTCGTTTGTATGATGAG 2194
OY 256 ALAHSTHR-----SEPTYRASPILARGSERLEUALASNALAASPLYSALA 272
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2195 GCCCACTCATTTGTGCAGAGAAACGTCCCGGTCTTAATTCACCTAAGAGAAATCA 2254
OY 273 ASN-----PROGLUVALSERTYRVALSERLEUYSERLEUALATYRPE 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2255 AGCAGCTGTTCCATACCTACCTACACACAGATATTTG-----TACTTC 2296
OY 289 METVALALAPROTHLEUCYSTYRGLNPROSERTYRPROARGSERALACYSILEARGYS 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2297 TTATTTGCTCTACCTTATCTACGCTGACAGCTATCCAGAGATCCACTGATACAG 2356
OY 309 GLYTRPVALLAARGGLNPHLEALALYSLEUVALLEPHERTHGLYPHEMETGLYPHEILE 328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2357 GGTATNGTGGCATAGATGATTCGACAGGCTTTGCTGCTTTTCTATGATGATGATC 2416
OY 329 ILEGLUGLYTRILEASPROILEVALARGANSERLYS---HISPROLEUYSGLYASP 347
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2417 TTTGAAGGCTTTGTGCCCCCTGTTTCCGAAATATCAACAGAGCCCTTACCGCT 2473
OY 348 LEULETYRALALILEGLUARGVALLEUYSLEUSERVAL-----PROASN 362
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2474 -----CGTGTCTGCTCTATGATGATTTTACTCCATCTTGCACGCT 2515
OY 363 LEUTYRVALTRPLEUCYSMETPHERTYRCSYPHEPHISLEUETHRPLEUANSILEUALA 382
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2516 GTGCTGATCTCTCTCTACTTTTGTGCTTTTGCACGTGCGTCAATGCTTGTCT 2575
OY 383 GLULEULEUCYPHEGLYASPARGLUPHETRYRGLYASPTTRTPASNALALYSERVAL 402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2576 GAGATGTTACGCTTTGTGACAGAGATGCTATAGAGATGGGGAATCCACAGTCATAC 2635
OY 403 GLYASPTYRTRARGMETRPARSMETPROVALHLSYSTRIMETVALARGHLSILEYR 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2636 TCCAACTATATATAGAACCTGGAATGCTGCTCATGACTGGCTATATTAATGCTTAC 2695
OY 423 -----PHEPROCYLSLEUARGSERLYSILEPROLYSTRHLEUALLEILELAPHE 440
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2696 AAGGACTTCTCTGCTGTTTCTCCAGAGATTCATCAATCTGCGGCATGTTAGCTGCTT 2755
OY 441 LEUVALSERALAPHEHLSGLULEUCYSILEALVALPROCYARG----- 456
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2756 GCTGATCTGCTGATATACAGAAATGCTGCTGCTGCTT-----TCTTGAAGCTTTTCTAT 2812
OY 457 -----LEUPLYSLEUTRIPALPHELEUGLYLMEPHEGLINVALPROLEUVALPHE 474
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2813 CCCGCTCTTGTGCTCTTCAATGTTCTTGGAAAGCTTTTAC-----TTC 2860
OY 475 ILEHRASNTYRLEUNGGLNUGARGPHEGLYSERTHVALGLYASNMELIPEPHERP 493
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2861 ATTGCAATAGATAGTCGAAAAAG-----CCGATTTGGAATGTTCTGATGCTGACT 2911
OY 494 -----PHEILEPHECYSLIPEHGLYGLN 501
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2912 TCTCTTTCTTGGCAATGAGACTTACTGCTTTATTTATTCCAA 2956
  
```

RESULT 11
 US-08-509-187D-3
 Sequence 3, Application US/08509187D
 Patent No. 5834283
 GENERAL INFORMATION:

```

Db 879 AAG---GACATGACTACTACAGCATATCAGAGCCCTCTGAACTGGCGCTCCCAAT 935
Oy 363 LeuTyValTrpLeuCysMetPheTyrcysPhePheHisLeuTrpLeuAsnIleLeuAla 382
Db 936 CACCTCATCTGCTGCTATCTCTCTACTGAGCTCTCCACCTCCGCTGAAATGCGCTGCT 995
Oy 383 GluLeuLeuCysPheGlyAspArgGluPheTyrcysAspTrpTrpAsnAlaLysSerVal 402
Db 996 GAGCTCATGAGATGTGGAGACGGGAGTTCTACCGGACTGCTGTGAACTCCGAGCTCTGC 1055
Oy 403 GlyAspTyrcysPheMetTrpAsnMetProValHisLysTrpMetValArgHisIleTyrc 422
Db 1056 ACCACTCTCTGCGACAAGTGGAAACATCCCTCTGCGACAAGTGGTGCATAGACACTTTCAC 1115
Oy 423 PheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPheLeuVal 442
Db 1116 AAGCCCATGCTTCTGCGAGCGGAGCAGCAGCATGTGATGGCAGACAGGGGTCTCTGGCC 1175
Oy 443 SerAlaValPheHisGluLeuCysIleLeuAlaValProCysArgLeuPheLysLeuTrpAla 462
Db 1176 TCGACCTTCTTCCAGAGTACCTGCTGAGCGTCCCTCTGCGAATGTTCGGCTCTGGCGC 1235
Oy 463 PheLeuGlyIleMetPheGluValProLeuVal---PheIleThrAsnTyrcysLeuGlnGlu 481
Db 1236 TTCACGGGCGATGATGCGTCACATCCCACTGGCTGCTGCGCGCTTTTTCAGAGGC 1295
Oy 482 ArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyLys 501
Db 1296 AACTAT-----GGCAAGCAGCAGCTGTGG---CTGTGCTCATCATCGAGCAG 1340
Oy 502 ProMetCysValIleLeuTyrcysTrpHisAsp 511
Db 1341 CCAATAGCGCTCATGTATGATGTCACGAC 1370

```

RESULT 9

US-09-103-754A-3

Sequence 3, Application US/09103754A

Patent No. 6344548

GENERAL INFORMATION:

APPLICANT: Farese, Robert

APPLICANT: Cases, Sylvaine

APPLICANT: Smith, Steven

APPLICANT: Erickson, Sandra

TITLE OF INVENTION: Diacylglycerol O-acyltran

TITLE OF INVENTION: siferase

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESS: Bozicevic & Reed

STREET: 285 Hamilton Avenue, Suite 200

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94301

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/103.754A

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Field, Bret E

REGISTRATION NUMBER: 37,620

REFERENCE/DOCKET NUMBER: 6510-105P

TELECOMMUNICATION INFORMATION:

TELEPHONE: 650 327 3400

TELEFAX: 650 327 3231

TELEX:

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-103-754A-3

```

Alignment Scores:

Pred. No.:	1.49e-62	Length:	629
Score:	652.50	Matches:	131
Percent Similarity:	87.58%	Conservative:	3
Best Local Similarity:	85.62%	Mismatches:	10
Query Match:	23.55%	Indels:	9
DB:	4	Gaps:	1

US-09-623-514A-2 (1-520) x US-09-103-754A-3 (1-629)

```

Oy 304 AlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIlePheThrGly 323
Db 2 GCATGTATACGGAAGGTTGGTGGCTGCTGCAATTGCAAAAGTGCATATTCACCGGA 61
Oy 324 PheMetGlyPheIleIleGlnGlnTrpIleAsnProIleValArgAsnSerLysHisPro 343
Db 62 TTCATGGGATTTATTAATGAAACAATATTAATCTATTTGTCAGAACTCAAGCATCT 121
Oy 344 LeuLysGlyAspLeuLeuTyrcysAlaIleGlnArgValLeuLysLeuSerValProAsnLeu 363
Db 122 TTGAAGGCGCATCTTCTATGCTATGTAAGAGAGTGTTGAAGCTTTCGCAAAATTTA 181
Oy 364 TyrValTrpLeuCysMetPheTyrcysPhePheHisLeuTrpLeuAsnIleLeuAlaGlu 383
Db 182 TATGTGTGCTGCTGCATGTCTACTGCTTCTCCACTTGTGGTAAACATATTGGCAGAG 241
Oy 384 LeuLeuCysPheGlyAspArgGluPheTyrcysAspTrpTrpAsnAlaLysSerValGly 403
Db 242 CTTCTGCTTCTGGGATGTAATTCACAAAGATGTGGAATGCAAAAGTGTGGGA 301
Oy 404 AspTyrcysArgMetTrpAsnMetProValHisLysTrp--MetValArgHisIleTyrc 422
Db 302 GATTAATGCGAAGATGTAATGTAATGCTGCTGCATTAAGGATGGGTCCGACATATATAC 361
Oy 423 Phe-ProCysLeuArgSerLysIle-ProLysThrLeuAlaIleIleAlaPheLeu 441
Db 362 CTTCCCGCTGCTTGGCGACAAGATATACCAAGACACCCGGCCATTAAACATTGGCTT 421
Oy 442 ValSerAla-----ValPheHis 447
Db 422 TCCCAAGCCCGCTGGAGGCTTTCCAT 448

```

RESULT 10

US-08-121-057-3

Sequence 3, Application US/08121057

Patent No. 5484727

GENERAL INFORMATION:

APPLICANT: CHANG, YA-YUAN

APPLICANT: CHANG, CATHERINE C. Y.

TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL

TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESS: LAHIVE & COCKFIELD

STREET: 60 STATE STREET, SUITE 510

CITY: BOSTON

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC Compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: ASCII Text


```

Db      444 AAGGACCCCTACAGCTGGCTGGCCCAAGCTTGATCATTCATGCATTCATATCTTATGTG 503
Oy      191 ALAlaIaPheThValGluLysLeuValLeuGlnLysTyrLleSerGluProValIle 210
Db      504 GCTACATTCATGATTTAGAGAACCGCTGTCCAGTGGGTGCCCGACAGACATAGAGGGCTG 563
Oy      211 PheLeuHisiIleIleThMetThrGluValLeuTyrProValTyrValThrLeuArg 230
Db      564 CTGCTCAATGTGGTTAACCTGGCCCAATTTATCTGCTTCCAGACAGCTGTGGCCTTACTG 623
Oy      231 CysAspSerAlaPheLeuSerGlyValThrLeuMetLeuThr---CysIleValTyrP 249
Db      624 GTTAGTGTATACATCTCCAGTGGGTCCCTGTGGTGTGGATCATATCATCATATCTTC 683
Oy      250 LeuLysLeuValSerTyrAlaHisiThSerTyr-----AspIleArgSer 264
Db      684 CTCAAGCTTTTCTCTACCGGAGATGTCAATCTGTGTGCCCCGACGAGGAGGTCAAGGCC 743
Oy      285 LeuAlaAsnAlaIaAspLys-----AlaAsnProGluValSerTyr 278
Db      744 AAAGCTGTGTCTGCAGGGAGAAAGTCACTGGGGCTGCTGCCAGAACATGTAAAGTAT 803
Oy      279 -----TyrValSerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyr 296
Db      804 CCGAGAACACCGAGACTACCGAGATCTGATTTACTTACATCTTGTCTCTCACTTGTGTAT 863
Oy      297 GluProSerTyrProArgSerAlaCysLysIleArgLysGlyTyrPvalAlaArgGlnPheAla 316
Db      864 GAACCAACATTTCTCCGATGCCCGCCGAAATACGAAAGCCCTTCTGCTACGGGGTCTT 923
Oy      317 LysLeuValIlePheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIle 336
Db      924 GAGATGCTCTTTTTCACCCAGCTTCACACTGGGGCTGATCCAGACAGATGATGGCTTACT 983
Oy      337 ValArgAsnSerLysHisProLeuLysGlyAspLeuLeuTyrAla-----IleGluArg 354
Db      984 ATCCAGAACTCCATGAAGACCTTCAAG--GACATGACATTTACAGAAATCATTTAGAGGT 1040
Oy      355 ValLeuLysLeuSerValProAsnLeuTyrValTyrLeuCysMetPheTyrCysPhePhe 374
Db      1041 CTTTAAAGCTGGGGGCTCCCAACCATCTGATATAGGCCATCTTCTATTTGGCTTTTC 1100
Oy      375 HisLeuTyrPheAsnAlaIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLys 394
Db      1101 CACTCATGTCTCAATGCTGTGGCAGAGCTCTTGCAAGTTTGGAGACCGGAGGTTCTACAGG 1160
Oy      395 AspTrrTyrAsnAlaLysSerValGlyAspTrrTyrArgMetTyrPasnMetProValHis 414
Db      1161 GACTGGGTGATGCTGAGTGTGTCAACCTACTTTTGGCAGAACTGGAAATATCCCGTGCAC 1220
Oy      415 LysTrrPmetValArgHisIleTyrPheProCysLeuArgSerLysIleProLysThrLeu 434
Db      1221 AAGTGGTGCATACAGACACTTCTAAAGCCATTCGTCTACAGCTGGCAGACAAACAAATGATG 1280
Oy      435 AlaIleIleIleIlePheLeuValSerAlaValAlaPheHisiGluLeuCysIleAlaValPro 454
Db      1281 GCCAGAGACTGGGGTCTTTTGGCTGCAGACCTTCTTCATGTAGTACCTAGTAGAGATTCCC 1340
Oy      455 CysArgLeuPheLysLeuTrrAlaPheLeuLysIleMetPheGlnValProLeuValPhe 474
Db      1341 CTGAGGATGTTCCCGCTCTGGGCAATTCACAGCAATGATGGTCCAGTCCACCTGGCTGG 1400
Oy      475 IleThrAsnTyrLeuGlnGluArgPheGlySerThrValGlyAsnMetIlePheTrrPhe 494
Db      1401 ATTGTGAACCCGCTTC-----TTCCAAAGGAACATATGCAATGACAGCTGTGTGG--- 1448
Oy      495 IlePheCysIlePheGlyGlnProMetCysValLeuLeuTyrTyrHisAsp 511
Db      1449 GTGACACTCATCATTTGGGCAACCGGTGGTGTGCTCATGTATGTCCACGAC 1499

```

[illegible]

Db 423 GCAGGTTATTTTACAGAACTTATCAGATGATGCTGTG--GATCCTATCCAG 479
 Qy 168 PheSerSerArgSerLeuArgAsp-----TrrProLeuPheMetCysLysLeu 184
 Db 480 GTGGTCTCTGTTTGAAGACCCCTACAGCTGGCTGCCCATCCGATATTATGCA 539
 Qy 185 LeuSerLlePheProLeuAlaAlaPheThrValGluLysLeuValLeuGlnLysTyrIle 204
 Db 540 TCCAAATTTTGTGTGGCTGCATTTTCAGATTGAGAAAGCGCTGGCAGATGGGCTG 599
 Qy 205 SerGluProValValLlePheLeuHisLleLleLlePheThrGluValLeuTyrPro 224
 Db 600 ACAGAGAGATGGGGCTGCTGCTACATGCTGTTACCTGGCCACAAATCATTTGCTTCCA 659
 Qy 225 ValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeu 244
 Db 660 GCAGCTGTGGCTTACGTGTGAGTGTATCATCCATCCAGTGGTCCGTTGCTGTGCA 719
 Qy 245 Thr---CysLleValTyrLeuLysLeuValSerTyrAlaHisThrSerTyrAspIleArg 263
 Db 720 TCATATCCATCATGTGCTCCCAAGCTTATTCCTACCGGATGTCACACTGTGGTGGCG 779
 Qy 264 -----SerLeuAlaAlaAlaAsp 270
 Db 780 CAGCGAAGGTCAAGCCAAAGCTGTCTCATCAGAGGAAGGACACTGGGGCTGCT-- 836
 Qy 271 LysAlaAsnProGluValSerTyr-----TyrValSerLeuLysSerLeuAlaTyrPhe 288
 Db 837 ---GCCAGCAAGCTGTGAGCTATCCAGAACCTGACCTGACGAGATCTCTATTTACTTC 893
 Qy 289 MetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysLleArgLys 308
 Db 894 ATCTTGCTCTCTACTTGTGTATGAACTCAACTTCTGCTGGTCCCGCAATACGAAG 953
 Qy 309 GlyTyrValAlaArgGlnPheAlaLysLeuValLlePheThrGlyPheMetGlyPheIle 328
 Db 954 CGCTTCTGCTACAGACAGTCTTTCAGATCTCTTTTTCACCACTTCAAGTGGGGCTG 1013
 Qy 329 IleGluGlnTyrLleAsnProLleValArgAsnSerLysHisProLeuLysGlyAspLeu 348
 Db 1014 ATCCAAAGATGATGGCTTACTATCCACACACTCCATGAAAGCCCTTCAG--GATATG 1070
 Qy 349 LeuTyrAla-----IleGluArgValLeuLysLeuSerValProAsnLeuTyrValTrr 366
 Db 1071 GACTATTCACGATCATATGAGCGCTCTTAAAGCTGGCGTCCCAACCACTTCGATCTGG 1130
 Qy 367 LeuCysMetPheTyrCysPhePheHisLeuTrrPheAsnLleLeuAlaGluLeuLeuCys 386
 Db 1131 CTATATCTTCTTCTATTTGTTTTCACCTCTGCTCTCATGCTGTGGCAGAGCTTCTGCA 1190
 Qy 387 PheGlyAspArgGluPheTyrLysAspTrrPheAsnAlaLysSerValGlyAspTyrTrr 406
 Db 1191 TTGGAGACCGCGAGTCTTACAGAGATTTGGATGCTGAGTCTGCTGCTACTTTGG 1250
 Qy 407 ArgMetTrrPheAsnMetProValHisLysTrrMetValArgHisLleTyrPheProCysLeu 426
 Db 1251 CAAACACTGGAATATCCCGGTCACAACTGTGATCAGACACTTCTCAAGACCTATCTC 1310
 Qy 427 ArgSerLysLleProLysThrLeuAlaLleLleLlePheLeuValSerAlaValPhe 446
 Db 1311 AGACATGCGCAGCAAAATGGGTGGCCAGACAGAGATATTTTGGACCTTACTTCTTC 1370
 Qy 447 HisGluLeuCysLleAlaValProCysArgLeuPheLysLeuTrrPheAlaPheGlyIle 466
 Db 1371 CARGAGTACCTAGTACAGCTGTTTTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1430
 Qy 467 MetPheGlnValProLeuValPheLleThrAsnTyrLeuGlnGluArgPheGlySerThr 486
 Db 1431 ATGGCTTACAGTCCACAGCTGTGATTTGGTGGCCGATTC-----TTCCAGAGGAC 1481
 Qy 487 ValGlyAsnMetLlePheTrrPheLlePheCysLlePheGlyGlnProMetCysValLeu 506
 Db 1482 TATGGCAATCAGCTGTGTG--GTGACACTCATCATTTGGGCAACCGGTGGCTGTGCTC 1538

Qy 507 LeuTyrTrrHisAsp 511
 Db 1539 ATGATGTCCACGAC 1553
 RESULT 6
 US-09-326-203A-15
 : Sequence 15, Application US/09326203A
 : Patent No. 6444876
 : GENERAL INFORMATION:
 : APPLICANT: Lasser, Mike
 : APPLICANT: Kuzlinsky, Diane
 : TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic
 : FILE REFERENCE: 17045/00/MO
 : CURRENT APPLICATION NUMBER: US/09/326, 203A
 : PRIOR APPLICATION NUMBER: 60/088, 143
 : PRIOR FILING DATE: 1998-06-05
 : PRIOR APPLICATION NUMBER: 60/108, 389
 : PRIOR FILING DATE: 1998-11-12
 : NUMBER OF SEQ. ID NOS: 46
 : SOFTWARE: PatentIn Ver. 2.1
 : SEQ. ID NO 15
 : LENGTH: 1766
 : TYPE: DNA
 : ORGANISM: Rattus sp.
 US-09-326-203A-15
 Alignment Scores:
 Pred. No.: 3,79e-75 Length: 1766
 Score: 773.50 Matches: 193
 Percent Similarity: 53.58% Conservative: 84
 Best Local Similarity: 37.33% Mismatches: 171
 Query Match: 27.91% Indels: 69
 DB: 4 Gaps: 17
 US-09-623-514a-2 (1-520) x US-09-326-203A-15 (1-1766)
 Qy 14 GluAsnGlyGlyGluPheValAspLeuAspArgLeuArgArgLysSerArgSer 33
 Db 99 CAGGCTGTGATGGGCCCATGCTAGACGAAGAGAGGTGGCAGACGCCCTGTGGCCCC 158
 Qy 34 AspSerSerAsnGlyLeuLeuLeuSerGlySerAspAsnAsnSerProSerAspAspVal 53
 Db 159 GAC-----TTGGGGCGCGGGGGGGGAGACCTCGGCTCGGCT--CCGGT 200
 Qy 54 GlyAlaProAlaAspValArgAspArgLleAspSerValAlaAsnAspAlaGlnGly 73
 Db 201 CCGGCTCCACCCACACACCCTGGGACAA-----GACCGGCGAGACC 239
 Qy 74 ThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyAspAsnAsnGlyGlyArgGly 93
 Db 240 AGCGTG-----GGCGAGCGC 254
 Qy 94 GlyGlyGluGlyArgGlyAsnAlaAspAlaThrPheThrTyrArgProSerValProAla 113
 Db 255 CACTGGAGCTGAGG-----TGC 272
 Qy 114 HisArgArgAlaArgGluSerProLeuSerSerAspAlaLlePheLysGlnSerHisAla 133
 Db 273 CAT---CGTTCGACAGCTTGTGTCAGCTCAGACAGCGGTTTC--ACCAATTACCGT 326
 Qy 134 GlyLeuPheAsnLeuCysValValValLleLleLleAlaValAsnSerArgLeuLleGlu 153
 Db 327 GGTATCTGTAATTTGGTGGGTGTGATCTGATCTGATGATGCAAGTTATTTTATGAG 386
 Qy 154 AsnLeuMetLysTyrGlyTrrPheLleArgThrAspPheTrrPheSerSerArgSerLeu 173
 Db 387 AATCTTATCAAGATGATGCAATCTGTG--GATCCCATCCAGAGGTGTCTGTTCTG 443
 Qy 174 ArgAsp-----TrrProLeuPheMetCysLysSerLeuSerLlePheProLeu 190

```

Db      820 GGGCTCCCTGTCGGCGTGATGGCGACACACACATCTTCTTCACAGCTCTTCTCTACCG 879
Qy      256 AhtshnrSerTyraPileArgSer-----LeuAlaAsnAl 268
Db      880 C-----GAGCTCAACTCATGTGTCCGACAGGCGCAGGCGCAAGGCTCTCTGCG 927
Qy      268 AAlaAspLysAlaAsnPro-----GluValSerTy-----TyrAlaIse 281
Db      928 AGGGAAGAAAGCCAGAGTGTCTGCTCCCGCACACCGGTGAGCTTACCCGACAAATGTGAC 987
Qy      281 rLeuYsSerLeuAlaLysPheMetValAlaProThrLeuCySTyGlnProSerTyrrp 301
Db      988 CTACCGCGAATCTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1047
Qy      301 oArgSerAlaCySIlArgLysGlyTrrpAlaAlaArgGlnPheAlaLysLeuValIlePh 321
Db      1048 CCGCTCTCCCGCTCCGAGAAAGGCGCTTGTGCTGCGACAGGAAATCTGAGATGCTGTCTT 1107
Qy      321 ePhrGlnPheMetGlyPheIleIleGlnIlnTyrrIleAsnProIleValArgAsnSerTy 341
Db      1108 CACCCAGCTCCAGAGTGGGCTGTATCCAGAGTGTGTGTCGCCACCATCCAGAAATCCCAT 1167
Qy      341 sHisProLeuLysGlyAspLeuLeuTyrrAla-----IleGluArgValLeuLysIse 359
Db      1168 GAAGCCCTTCAG--GACATGAGCTACTACAGCATCATCATGAGCGGCTCTCTGAGCTGGC 1224
Qy      359 rValProAsnLeuTyrrValTrrLeuCySmetPheTyrrCySPhenHisLeuTrrLeuAs 379
Db      1225 GGTCCCAATTCACCTCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 1284
Qy      379 nIleLeuAlaGlnLeuLeuCySPhenGlyAspArgGlnPheTyrrLysAspTrrTrpAsnAl 399
Db      1285 TGCCGTGGTGTGAGCTATGAGTTGGAAACCGGGGATCTTACCGGGAGCTGGTGAATCTC 1344
Qy      399 aLysSerValGlyAspTrrTrpArgMetTrrAsnMetProValHisLysTrrMetValAr 419
Db      1345 CAGATCTCTCAGCTACTTCTGTGCGAAGAACTGGAACATCCTGTGCAAAAGTGGTGCATCAG 1404
Qy      419 gHisIleTyrrPheProCySLeuArgSerLysIleProLysTrrIleAlaIleIleIleAl 439
Db      1405 ACACCTCTCTACAAAGCATGCTTCGACGGGCGACAGCAGCAAGTGTGATGCCAGACAGGGGT 1464
Qy      439 aPheLeuValSerAlaValPheHisGlnLeuCySIlleAlaValProCyArgLeuPheLys 459
Db      1465 GTTCCTGGCGCTGGCTTCTCTCTCAGAGTACCTGTGAGCGTCCCTCTCGAATGTTCGG 1524
Qy      459 sLeuTrrAlaPheLeuGlyIleMetPheGlnValProLeuVal---PheIleThrAsnTy 478
Db      1525 CCTCTGGGCTTTCACGGGCAATATGCTCAGATCCCATCTGGCTGTGTGCGCGCGCTT 1584
Qy      478 rLeuGlnGlnArgPheGlySerThrValGlyAsnMetIlePheTrrPheIlePheCySIl 498
Db      1585 TTTCCAGGGCGCAACTAT-----GGCAACGAGAGCTGTGTG---CTGCGGCTCAT 1629
Qy      498 ePheGlyGlnProMetCySValLeuLeuTyrrTyrrHisAsp 511
Db      1630 CATCGGACAGCAATATGCGCTCTCATGTACGTCCAGAC 1669

```

RESULT 5
US-09-103-754A-2
Sequence 2, Application US/09103754A
Patent No. 6344548

GENERAL INFORMATION:
APPLICANT: Farese, Robert
APPLICANT: Cases, Sylvaine
APPLICANT: Smith, Steven
APPLICANT: Erickson, Sandra
TITLE OF INVENTION: Diacylglycerol O-acyltran
TITLE OF INVENTION: sferase
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed
STREET: 285 Hamilton Avenue, Suite 200

```

? CITY: Palo Alto
? STATE: CA
? COUNTRY: USA
? ZIP: 94301
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? OPERATING SYSTEM: DOS
? SOFTWARE: FASTSEQ for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/09/103,754A
? FILING DATE:
? CLASSIFICATION:
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? ATTORNEY/AGENT INFORMATION:
? NAME: Field, Bret E
? REGISTRATION NUMBER: 37,620
? REFERENCE/DOCKET NUMBER: 6510-105P
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 650 327 3400
? TELEFAX: 650 327 3231
? TELEX:
? INFORMATION FOR SEQ ID NO: 2:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1650 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? US-09-103-754A-2

Alignment Scores:
Pred. No.: 3,39e-75 Length: 1650
Score: 773.50 Matches: 191
Percent Similarity: 51.818 Conservative: 81
Best Local Similarity: 36.388 Mismatches: 178
Query Match: 27.918 Indels: 75
DB: 4 Gaps: 15

US-09-623-514A-2 (1-520) x US-09-103-754A-2 (1-1650)
Qy      8 GlyValThrThrValThrGlnAsnGlyGlyGlnPheValAspLeuAspArgLeuArg 27
Db      141 GGCTCGGGGTTCCTCCGTCCAGAGGTGTAGTGGCCCAAGTGAAGAGCAGAGGTGCGCA 200
Qy      28 ArgArgLysSerArgSerAspSerAsnGlyLeuLeuLeuSerCylSerAspAsnAsn 47
Db      201 GACGCGGCTGTGAGCCCGAC-----TTGGGCGCGCGGGGTGACGCGCGCG 245
Qy      48 SerProSerAspAspValGlyAlaProAlaAspValArgAspArgIleAspSerValVal 67
Db      246 GCTCGCGCTCCG-----GCTCCAGCCCATACCCGCGACAAA----- 281
Qy      68 AsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyLysAsn 87
Db      281 ----- 281
Qy      88 AsnGlyGlyGlyArgGlyGlyGlyGlnGlyArgGlyAsnAlaAspAlaThrPheThrTyrr 107
Db      282 ---GACGGCGGACACAGCGTGGCGCGAGCTCTACTGGAGTCTAGAGTGC----- 326
Qy      108 ArgProSerValProAlaHisArgAlaArgGluSerProLeuSerSerAspAlaIle 127
Db      327 -----CATGCTCTGCAAGATTTCTTTGTTCACTCAGCTCAGACAGTGTGT 365
Qy      128 PheLysGlnSerHisAlaGlyLeuPheAsnLeuCySValAlaValLeuIleAlaValAsn 147
Db      366 TTC---AGCAATATAGTGTATCTCGAATGTGTGTGTGTGTATGTATCTCTGATAT 422
Qy      148 SerArgLeuIleIleGlnAsnLeuMetLysTyrrGlyTrrPheLeuIleArgThrAspPheTrrp 167

```

```

Db 181 CAGTTATTTACACCATGACGAGCTTGTATTCACGCTATGTCACCTAAGGTGAT 240
Oy 233 SerAlaPheLeuSerGlyValThrLeuMetLeuThrCysIleValITrPLeuLysLeu 252
    |||||
Db 241 TCTGCCCTTATACAGGTACAGCTTGATCTCTCTCATGCTATGCTGCTCAATG 300
Oy 253 ValSerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspLysAla 272
    |||||
Db 301 GTTCTTACCTCTACTACTACTATGACATAGAACCCCTACTATATCATGATGATAAGGCC 360
Oy 273 AspProGluValSerTyrTyrValSerLeuLysSerLeuAlaTyrPheMetValAlaPro 292
    |||||
Db 361 AATCTGTAAGTCTCTACTATGATGCTTAAGAGCTTGCTTATTTTCATGCTGCTGCC 420
Oy 293 ThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArgLysGlyTyrValAla 312
    |||||
Db 421 ACATGGTATTATGACGACGACGATATCCAGTTCATGATTCGGAAGGGTGGGTGGCT 480
Oy 313 ArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleIleGlnTyr 332
    |||||
Db 481 CGTCAATTTTCAAAAGCTGATATTCACGAGCTCATGGAATTTATATAGACCAATAT 540
Oy 333 IleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeuTyrAlaIle 352
    |||||
Db 541 ATTAATCTTATTTGATGAACTCAAGACATCTCTGAAAGGGGACCTGTATATGCTATT 600
Oy 353 GluArgValLeuLysLeuSerValProAsnLeuTyrValITrPLeuCysMetPheTyrCys 372
    |||||
Db 601 GAAAGAGCTGTGAAGCTTTCAGTTCCAATCTATATGTCGCTCAGCTCATGTTCTACTGC 660
Oy 373 PhePheHisLeuTrpLeuAlaIleLeuAlaGluLeuLysCysPheGlyAspArgGluPhe 392
    |||||
Db 661 TTCTTCCACCTTTGGTTAAACATATGCGAGAGCTCTGCTGCGGGGACCGGAATTC 720
Oy 393 TyrLysAspTrpTrpAsnAlaLysSerValGlyAspTyrTrpArgMetTrpAsnMetPro 412
    |||||
Db 721 TACAAGATTTGGTGATGCAAAAGCGTTGGAGATTATGGAATATGGAATATGCT 780
Oy 413 ValHisLysTrpMetValArgHisLysIleTyrPheProCysLeuArgSerLysIleProLys 432
    |||||
Db 781 GTTCACAAATGATGATGTCACATGATGATCTTCCGCGCCGACATCAAGATTCACAAA 840
Oy 433 ThrLeuAlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLysCysIleAla 452
    |||||
Db 841 GTCACCCGCAATTCATGCTTCTTCTTATGCTGACATCTTTCATGATGATGATGCA 900
Oy 453 ValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetPheGlnValProLeu 472
    |||||
Db 901 GTTCCTTGCGGCTCTCTCATATGAGGCTTTCATGGAATATATGTTTCAGGTCCTTG 960
Oy 473 ValPheIleThrAsnTyrLeuGlnArgPheGlySerThrValGlyAsnMetIlePhe 492
    |||||
Db 961 GTCCTTATCAAAACCTTTTACAAAGAGTTGGCTCCATGCGGGAACATCATGCTTT 1020
Oy 493 TrpPheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTyrTyrHisAspLeu 512
    |||||
Db 1021 GGTTCACGCTTCTGCAATTTGGACCAACGATGTCGCTTATATACATGACGCTG 1080
Oy 513 MetAsnArgLysGlySerMetSer 520
    |||||
Db 1081 ATGACCGCAAGATCATGCTCC 1104

```

```

; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1976
; TYPE: DNA
; ORGANISM: Yeast
; US-09-165-042-2

Alignment Scores:
Pred. No.: 2,51e-77 Length: 1976
Score: 794.00 Matches: 205
Percent Similarity: 52.53% Conservative: 86
Best Local Similarity: 37.00% Mismatches: 190
Query Match: 28.65% Indels: 73
DB: 3 Gaps: 20

US-09-623-514a-2 (1-520) x US-09-165-042-2 (1-1976)
Oy 4 LeuAspSerAlaGly-----ValThrThrValThrGluAsnGlyGlyGluPheVal 21
    |||||
Db 91 GTTGAACGCGCTGTGTGAGGCGGTACCCGCGGCTACGCGCGGCGGACGAGGCGCATGCGCG 150
Oy 22 AspLeuAspArgLeuAlaArg-----ArgLysSerArgSerAspSerSer 36
    |||||
Db 151 CCGTTGCTCTAGGCGCCGAGAGTGGGCGCGCGCTCGGCGGCTACGACACCGGCGAGGCC 210
Oy 37 AsnGlyLeuLeuLeuSerGlySer----- 44
    |||||
Db 211 CACGCTTGCTGCGCGCGGCTGCGGCTGAGGCCATGAGCCGCGGACCTCCGCGCG 270
Oy 45 -----Asp-AsnAsnSerProSerAsp---AspValGlyAlaProAlaAspValArgAs 61
    |||||
Db 271 CCGAGAGACAGAGGTGCGCGCCCTCGAGCACGCGGCGGCGGCGCTCGGCGGCGGAGA 330
Oy 61 PArgIleAspSerValValAsnAspAlaGlnGly---ThrAlaAsnLeuAlaGlyAs 80
    |||||
Db 331 AGAG-----GTGCGGAGCGCGCTGCGGCGCGGACGCTGGAGCCGCGGGGGA 378
Oy 80 PAsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGluArgGlyAs 100
    |||||
Db 379 CGCGCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 438
Oy 100 nAlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgArgAlaArgGluSe 120
    |||||
Db 439 GCTGAGGTGC-----CATGCGCTCGAGATTC 465
Oy 120 rProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVa 140
    |||||
Db 466 TTATTCAGCTCTGACAGTGGCTTC---ACGACACTACCGTGGCATCTGTAAGTGTGT 522
Oy 140 lValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTr 160
    |||||
Db 523 GGTGATCTATATCTTGACATATGCGCGGTATTTCTGGAAGACTCATCAAGATGCGAT 582
Oy 160 PleuIleArgThrAspPheTrpPheSerArgSerLeuArgAsp-----TrpTr 177
    |||||
Db 583 CTGCTG---GACCCCAATCCAGGCTGCTCTGCTTCTGTAAGAGATCCCATGATGCTGCC 639
Oy 177 OleuPheMetCysLysLeuSerLeuSerIlePheProLeuAlaAlaPheThrValGly 197
    |||||
Db 640 CGGCCCATGCTGCTGTTATGCGGCAATGCTTGTGCTGCTGATTCACAGTGTGAGA 699
Oy 197 sLeuValLeuGlnLysTyrIleSerGluProValValIlePheLeuHisIleIleThr 217
    |||||
Db 700 GCGCGCTGCGGTGGGTGCGGTGACGAGGAGCGGCGGACCTGCTGACGTAGCAACCT 759
Oy 217 rMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSe 237
    |||||
Db 760 GCGCACCATCTGTCGTTCCAGCGGCTGCGGTCTGCTGCTGATGAGTATCATCTCAGT 819
Oy 237 rGlyValThrLeuMetLeuThr---CysIleValITrPLeuLysLeuValSerTyrAl 256
    |||||

```

```

; RESULT 4
; US-09-165-042-2
; Sequence 2, Application US/09165042
; Patent No. 6100077
; GENERAL INFORMATION:
; APPLICANT: Sturley, Stephen L.
; APPLICANT: Oelkers, Peter
; TITLE OF INVENTION: ISOLATION OF A GENE ENCODING DIACYLGLYCEROL
; FILE REFERENCE: 0575/56331
; CURRENT APPLICATION NUMBER: US/09/165,042

```


Db 112 CTC-----TCCGATCTTCCCGTCGGACGATGTTGGACCTCGCGGCCGCA 159
 Qy 60 ArgAspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGly 79
 Db 160 AGGATCGGGTGGATTGATTCGCTCCGAGAGAGGCTCAGAGGAACACGAAATTAGCT 216
 Qy 80 AspaAsnGlyGlyGlyAspaAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGly 99
 Db 217 -----GGCGAGATCGCCGAACTAGGGAATCCGCGGAGC----- 252
 Qy 100 AsnAlaAspAlaThrPheThrArgProSerValProAlaHisArgAlaArgGlu 119
 Db 253 -----GATGTAAGGTTACGATCGACCGCTCGCTCCAGCTCATCGAGAGAGGAG 306
 Qy 120 SerProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuGly 139
 Db 307 ACTCCTCTCAGCTCGACGCTATCTTCAAAACAAAGCATGACGAGATTGTTCAACCTCTG 366
 Qy 140 ValValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGly 159
 Db 367 GATGTTGTTCTGTTGCTGTAAAGTAACTATCATCGAAACCTCATGAAGTATGGT 426
 Qy 160 TrpLeuIleArgThrAspPheTrpPheSerArgSerLeuArgAspTrpProLeuPhe 179
 Db 427 TGGTTGATCAGAACTGATTTTGGTTAGTTCTACATCTTACGAGACTGGCCGCTTTC 486
 Qy 180 MetCysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuVal 199
 Db 487 ATGTGTTGCTTCACTTTCGCTGCTTTCCTTGGCTGCTTCAACGCTGAGAAATGGTA 546
 Qy 200 LeuGlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleIleThr 219
 Db 547 CTTCGAAATTAATATGAGCGCTGTGCAATCTTCAATGATTAATCAACAGACA 606
 Qy 220 GluValLeuLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyVal 239
 Db 607 GAGGCTCTGTATCAGCTCAGCTCAGCTCAGCTGATGCTGCTTCTTGTGCTGCTGCT 666
 Qy 240 ThrLeuMetLeuLeuThrCysIleValTyrLeuLysLeuValSerTyrAlaHisThrSer 259
 Db 667 AGCTTGATGCTGCTCAGCTGATGCTGCTGAGATGCTGCTGCTGCTGCTGCTGCTGCT 726
 Qy 260 TyrAspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTyr 279
 Db 727 TAGGACATTAAGAACCTCGCCAAATCAGCTGATTAAGTCACTGAAATCTCCCACTAT 786
 Qy 280 ValSerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuLysTyrGlnProSer 299
 Db 787 GTTACTTGAAGAGCTGGCGCTATTTCAATGCTGCTCCACACTGCTGTTATCAACCAAGC 846
 Qy 300 TyrProArgSerAlaCysIleArgLysGlyTyrValAlaArgGlnPheAlaLysLeuVal 319
 Db 847 TATCCAGCTTCTTCATGATCGGAAAGGTTGGTGGCTCGTCAACTGCAAAAGCTGCTC 906
 Qy 320 IlePheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsn 339
 Db 907 AATATCAGTGAAGCTATGATTAATAGAGCATATATATATATATATATATATATATAT 966
 Qy 340 SerLysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSer 359
 Db 967 TCMAAGCATCTCTGAAGAGGACCTTCTATATGCTATGAAGAGATGTTGAAGCTTTTA 1026
 Qy 360 ValProAsnLeuTyrValTyrLeuLysMetPheTyrCysPhePheHisLeuTyrLeuAsn 379
 Db 1027 GTTCAAAATCTATATGTTGGTGTGATGCTCTACTGCTTCTTCACTTTGGTTAAAC 1086
 Qy 380 IleLeuAlaGluLeuLeuLysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAla 399
 Db 1087 AATATGGCAGAGCTCTCTGCTGCTGGGACCGTGAATCTTCAAAAGATGTTGAATGCA 1146
 Qy 400 LysSerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArg 419
 Db 1147 AAAAGCTTGGAGATTATTTGAGAAATGTGAATATGCTGTTCAACAAATGATGTTTCA 1206

Qy 420 HisIleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIleAla 439
 Db 1207 CATGTAATCTTCCGCTCGGCTGGCCATCAAGATACAAAAGTACCCTTATCATTTAGCT 1266
 Qy 440 PheLeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLys 459
 Db 1267 TTTCTTAGTCTCGCACTCTTTCATGATGATGATGATGATGATGATGATGATGATGAT 1326
 Qy 460 LeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeu 479
 Db 1327 CTAATGGCTTTCATGGAATTAATGTTCAAGCTCCCTTGTGCTTATCAACAACTTTTA 1386
 Qy 480 GlnGluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePhe 499
 Db 1387 CAAGAAAGGTTGGCTCATGATGGGGAACATGATCTTTGGTTCACTTCTGATTTTC 1446
 Qy 500 GlyGlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMet 519
 Db 1447 GGACAAACGATGTGTGGCTTCTTATATCAATGATGATGATGATGATGATGATGATG 1506
 Qy 520 Ser 520
 Db 1507 TCC 1509
 Db 1507 TCC 1509
 RESULT 3
 US-09-593-359-1
 ; Sequence 1, Application US/09593359
 ; Patent No. 6552250
 ; GENERAL INFORMATION:
 ; APPLICANT: Laroche, Andre J.
 ; APPLICANT: Nykiforuk, Cory L.
 ; APPLICANT: Weslake, Randall J.
 ; TITLE OF INVENTION: Diacylglycerol O-acetyltransferase
 ; FILE REFERENCE: 24015US0
 ; CURRENT APPLICATION NUMBER: US/09/593,359
 ; CURRENT FILING DATE: 2000-06-14
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 1
 ; LENGTH: 1446
 ; TYPE: DNA
 ; ORGANISM: Brassica napus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (82)..(1107)
 US-09-593-359-1
 Alignment Scores:
 Pred. No.: 4,43e-192 Length: 1446
 Score: 1833.00 Matches: 342
 Percent Similarity: 95.65 Conservative: 10
 Best Local Similarity: 92.938 Mismatches: 16
 Query Match: 66.158 Indels: 1
 Gaps: 0
 US-09-623-514A-2 (1-520) x US-09-593-359-1 (1-1446)
 Qy 153 GluAsnLeuMetLysTyrGlyTyrPheLeuIleArgThrAspPheTrpPheSerArgSer 172
 Db 2 GAATAATCTCATGAGACGCTGCTGATGACAGACGATTTCTGCTTATGTTCAACGCTCG 61
 Qy 173 LeuArgAspTrpProLeuPheMetCysLysIleSerLeuSerIlePheProLeuAlaAla 192
 Db 62 CTGCGAGATTG-CGGCTTTTCATGTTGTCTCTCCCTTCAATCTTTCCTTGGCTGGCC 120
 Qy 193 PheThrValGluLysLeuValLeuGlnLysTyrIleSerGluProValValIlePheLeu 212
 Db 121 TTACCGTGAAGAAATTAATGATCTTCAAGAAATGATGATGATGATGATGATGATGATG 180
 Qy 213 HisIleIleIleThrMetThrGluValLeuTyrProValTyrValThrLeuArgCysAsp 232


```

Db 237 ATGGCATTGTTGATTCCTGCTGCGCTACTACGGTGAACGAGAACGGTGGCGAGAGTTC 296
QY 21 ValAspLeuAspArgLeuArgArgArgSerArgSerArgSerArgSerArgSerArgSerArg 40
Db 297 GTGATCTTGATGATGGCTGCTGCGAGGAATGAGATTCGATTCCTTAACGAGACTTCTT 356
QY 41 LeuSerGlySerAspAsnAsnSerProSerAspPvalGlyAlaProAlaAspValArg 60
Db 357 CTCTCGGTTCCTCAATATATCTCTCGGTGATGATGTTGAGACTCCCGCGAGCTTAGG 416
QY 61 AspArgIleAspSerValIleAsnAspAspAlaGlyIleThrIleAsnLeuAlaGlyAsp 80
Db 417 GATCGGATGATTCCTGTTAAGATGACGCTGAGGAGACAGCCAAATTTGGCCGGAGAT 476
QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGly 100
Db 477 AATAACGGTGGTGGCGATATATACGGGTGGTGAAGAGCGCGGAGAGAGAGAGAGAAC 536
QY 101 AlaAspAlaThrPheThrTyraArgProSerValProAlaHisArgAlaArgGlySer 120
Db 537 GCCGATGCTACGCTTACGTACGACCGTGGTCCAGCTCAGCTCAGTGGAGGCGAGAGAGT 596
QY 121 ProLeuSerSerAspAlaIlePheIleGlnSerHisAlaGlyLeuPheAsnLeuGlyVal 140
Db 597 CCACCTAGCTCCGACGCAATCTTCAACAGAGCCATGCGGATTTCAACCTCTGTGTA 656
QY 141 ValIleAlaIleAlaValAsnSerArgLeuIleIleGlyAsnLeuMetIleTyrglyTrp 160
Db 657 GTAGTCTTATCTCTGTAACATAGCTATCATCATGAAATCTTATGAAATGATGGTGG 716
QY 161 LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180
Db 717 TTGATCAGAACGGATTCGTGGTTAGTTCAGATCCCTGCGAATTTGGCCGCTTTTCATG 776
QY 181 CysCysIleSerLeuSerIlePheProIleAlaIlePheThrValGlyIleLeuValIleu 200
Db 777 TGTGTGATATCCCTTTGATCTTCTCTTCTGCTGCTGCTTACGGTTGAGAAATGGTACT 836
QY 201 GlnIleTyrlleSerGlyProValIlePheLeuHisIleIleIleIleIleIleIleIle 220
Db 837 CAGAAATACATATCAACACCTGTCATCTTCTCATATATATATATATATATATATATAT 896
QY 221 ValLeuTrpProValIleValIleThrLeuArgCysAspSerIleAspLeuSerGlyValThr 240
Db 897 GTTTGTATCCAGTTTACGTCACCTGAAAGGTGATTCGCTTTTATATAGGTGTCACT 956
QY 241 LeuMetLeuLeuThrCysIleValIleTrpLeuIleValSerTyraIleHisIleTrpSerTy 260
Db 957 TTGATGCTCTCTCACTTGCATTTGCTGCTAAAGTTGGTTCTTATGCTCATACTAGCTAT 1016
QY 261 AspIleArgSerLeuAlaAsnAlaIleAspIleAspIleAsnProGlyValSerTyryVal 280
Db 1017 GACATTAAGATCCCTACCAATGACGCTGATTAAGCCCAATCTCTACTACTAGT 1076
QY 281 SerLeuIleSerLeuAlaIleTrpPheMetValAlaProThrIleuGlyIleProSerTy 300
Db 1077 AGCTTAAGAGCTTGGCATATTTTCATGCTGCTCCACATGCTGTTATCAGCAAGTAT 1136
QY 301 ProArgSerAlaCysIleArgGlyIleTrpValAlaArgGlnPheIleAlaIleValIle 320
Db 1137 CCACGTTTCCATGATACGGAAGGTTGGTGGCTGCTCAATTTGCCAAACCTGGTCATA 1196
QY 321 PheThrGlyPheMetGlyPheIleIleGlnGlnTyrlleAsnProIleValAlaArgSer 340
Db 1197 TTCACGGATTCATGGGATTTATATAGAACATATATATATATATATATATATATATAT 1256
QY 341 LysHisProLeuIleGlyAspLeuLeuTyraIleIleGlyValLeuIleuIleuSerVal 360
Db 1257 AACCATCCCTTGAAGCGCATCTTCAATATGCTATGAAAGAGTGTGAAGCTTTCAGT 1316
QY 361 ProAsnLeuTyraIleTrpLeuGlyMetPheTyraCysPhePheHisIleuTrpLeuAsnIle 380
Db 1317 CCAAATTTATATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1376

```

```

QY 381 LeuAlaGlyLeuLeuGlyCysPheGlyAspArgGlyPheTyraIleAspTrpTrpAsnAlaIle 400
Db 1377 TTGCGAGAGCTTCTCTGCTCGGGGATTCGGAATTCACAAATATGTTGATGCAAAA 1436
QY 401 SerValGlyAspTyraTrpArgMetTrpAsnMetProValHisIleTyraIleValArgHis 420
Db 1437 AGTGTGGAGATTTACGGAAGATGGAATATGCTGCTCATATATATATATATATATATAT 1496
QY 421 IleTyraPheProCysLeuArgSerIleProIleThrIleAlaIleIleIleIleAlaPhe 440
Db 1497 ATATACCTCCCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1556
QY 441 LeuValSerAlaValPheHisGlyLeuGlyIleAlaValProCysArgLeuPheIleGly 460
Db 1557 CTAGTCTCTCACTCTTCTGATGAGCATGATGATGATGATGATGATGATGATGATGATGAT 1616
QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyraIleGln 480
Db 1617 TGGCTTTTCTTCTGATATATATATATATATATATATATATATATATATATATATAT 1676
QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500
Db 1677 GAAAGGTTTGGCTCAACGCTGGGGAACATGATCTTGGTTCATCTTCTGATTTGCGA 1736
QY 501 GlnProMetCysValLeuLeuTyraIleHisAspLeuMetAsnArgIleGlySerMetSer 520
Db 1737 CAACCGATGTGTGCTGCTTATATATATATATATATATATATATATATATATATATAT 1796

RESULT 2
US-09-593-359-3
; Sequence 3, Application US/09593359
; Patent No. 6552250
; GENERAL INFORMATION:
; APPLICANT: Larocche, Andre J.
; APPLICANT: Nykliforuk, Cory L.
; APPLICANT: Weslake, Randall J.
; TITLE OF INVENTION: Diacylglycerol O-acyltransferase
; FILE REFERENCE: 24015US0
; CURRENT APPLICATION NUMBER: US/09/593,359
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1512
; TYPE: DNA
; ORGANISM: Brassica napus
; FEATURE:
; OTHER INFORMATION: DGAT1
; NAME/KEY: CDS
; LOCATION: (1)..(1512)
US-09-593-359-3

Alignment Scores:
Pred. No.: 1,43e-244 Length: 1512
Score: 2309.50 Matches: 442
Percent Similarity: 88.10% Conservative: 17
Best Local Similarity: 84.84% Mismatches: 43
Query Match: 83.35% Indels: 19
DB: 4 Gaps: 5

US-09-623-514a-2 (1-520) x US-09-593-359-3 (1-1512)
QY 1 MetAlaIleLeuAspSerAlaGlyValIleThr-----ValThrGlnAsnGlyGlyGly 19
Db 1 ATGGCATTGTTGATTCCTGAGCGCTGCTGACCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 51
QY 20 PheValAspLeuAspArgLeuArgArgGlySerArgSerArgSerArgSerArgSerArg 39
Db 52 GTGCGGATCTCGACAGGCTCCACGCTGTAATTCGAGTTCGATTCCTTCAACGAGACTC 111
QY 40 LeuLeuSerGlySerAspAsnAsnSerProSerAspPvalGlyAlaProAlaAspVal 59
Db 111

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 31, 2003, 05:06:49 ; Search time 97 Seconds
(without alignments)
2366.178 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771
Sequence: 1 MAILDSAGVTTVTENGGEF.....OPMCVLLYYHDLNKRKGSMS 520

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+pn.model -DEV=xlp
-O=/cg2.1/USPTO.spool/US09623514/runat.29082003.152200.102/app.query.fasta.1.711
-DB=Issued_Patents.NA -QFWT=fastap -SUFFIX=rrl -MINMATCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCCALIGN=200 -THR.SCORE=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09623514 @CGN.1.1.85.erna1.29082003.152200.102 -NCPD=6 -ICPD=3
-NO_MMAP -LARGEDEVTRY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOGS
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued_Patents.NA:*
1: /cg2.6/ptodata/1/ina/5A.COMB.seq:*
2: /cg2.6/ptodata/1/ina/5B.COMB.seq:*
3: /cg2.6/ptodata/1/ina/6A.COMB.seq:*
4: /cg2.6/ptodata/1/ina/6B.COMB.seq:*
5: /cg2.6/ptodata/1/ina/PTUS.COMB.seq:*
6: /cg2.6/ptodata/1/ina/Backfiles1.seq:*

-Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	1942	4	US-09-326-203A-1 Sequence 1, Appl1
2	2309.5	83.3	1512	4	US-09-593-359-3 Sequence 3, Appl1
3	1833	66.1	1446	4	US-09-593-359-1 Sequence 1, Appl1
4	794	28.7	1976	3	US-09-165-042-2 Sequence 2, Appl1
5	773.5	27.9	1650	4	US-09-103-754A-2 Sequence 2, Appl1
6	773.5	27.9	1766	4	US-09-326-203A-15 Sequence 15, Appl1
7	773.5	27.9	1766	4	US-09-326-203A-16 Sequence 16, Appl1
8	709.5	25.6	1895	4	US-09-326-203A-14 Sequence 14, Appl1
9	652.5	22.9	629	4	US-09-103-754A-3 Sequence 3, Appl1
10	412.5	14.9	4011	1	US-08-121-057-3 Sequence 3, Appl1
11	412.5	14.9	4011	2	US-08-509-187D-3 Sequence 3, Appl1
12	412.5	14.9	4011	2	US-09-121-396-3 Sequence 3, Appl1

13	412.5	14.9	4011	5	PCR-US93-09704A-3 Sequence 3, Appl1
14	405.5	14.6	4079	1	US-08-121-057-2 Sequence 2, Appl1
15	405.5	14.6	4079	2	US-08-509-187D-2 Sequence 2, Appl1
16	405.5	14.6	4079	5	US-09-121-396-2 Sequence 2, Appl1
17	405.5	14.6	4079	5	PCR-US93-09704A-2 Sequence 1, Appl1
18	375	13.5	1607	4	US-09-328-857A-1 Sequence 1, Appl1
19	364.5	13.2	1509	4	US-09-328-857A-2 Sequence 2, Appl1
20	364	13.1	2040	3	US-09-165-042-4 Sequence 4, Appl1
21	299	10.8	275	4	US-09-326-203A-5 Sequence 4, Appl1
22	284.5	10.3	253	4	US-09-326-203A-7 Sequence 7, Appl1
23	284	10.2	234	4	US-09-326-203A-3 Sequence 3, Appl1
24	277	10.0	267	4	US-09-326-203A-4 Sequence 4, Appl1
25	275	9.9	254	4	US-09-326-203A-8 Sequence 8, Appl1
26	275	9.9	254	4	US-09-313-294A-580 Sequence 580, App
27	266	9.6	262	4	US-09-326-203A-9 Sequence 9, Appl1
28	266	9.6	262	4	US-09-313-294A-1662 Sequence 1662, Ap
29	224.5	8.1	518	4	US-09-326-203A-12 Sequence 12, Appl
30	187.5	6.8	519	4	US-09-326-203A-11 Sequence 11, Appl
31	166.5	6.0	325	4	US-09-326-203A-10 Sequence 10, Appl
32	151	5.4	996	1	US-08-121-057-1 Sequence 1, Appl1
33	151	5.4	996	2	US-08-509-187D-1 Sequence 1, Appl1
34	151	5.4	996	2	US-09-121-396-1 Sequence 1, Appl1
35	151	5.4	996	5	PCR-US93-09704A-1 Sequence 1, Appl1
36	150	5.4	6755	3	US-08-931-999-4 Sequence 4, Appl1
37	144	5.2	1239	4	US-09-134-001C-2048 Sequence 2048, Ap
38	142.5	5.1	3967	4	US-09-672-749-14 Sequence 14, Appl
39	122.5	4.4	1230025	4	US-09-198-452A-1 Sequence 1, Appl1
40	120	4.3	3707	1	US-08-116-101A-1 Sequence 1, Appl1
41	117	4.2	7557	1	US-08-464-136-4 Sequence 4, Appl1
42	117	4.2	7557	2	US-08-349-131-4 Sequence 4, Appl1
43	117	4.2	7557	3	US-08-470-297A-4 Sequence 4, Appl1
44	117	4.2	7557	5	PCR-US91-07149-4 Sequence 4, Appl1
45	114	4.1	1029	4	US-09-205-815B-45 Sequence 45, Appl1

ALIGNMENTS

RESULT 1
US-09-326-203A-1
; Sequence 1, Application US/09326203A
; Patent No. 6444876
; GENERAL INFORMATION:
; APPLICANT: Ruzinsky, Mike
; TITLE OF INVENTION: Acid-CoA:Cholesterol Acyltransferase Related Nucleic
; FILE REFERENCE: 17045/00/MO
; CURRENT FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,143
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/108,389
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1942
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-326-203A-1

Alignment Scores:

Pred. No.: 2,916-295
Score: 2771.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 100.00%
DB: 4
Gaps: 0

US-09-623-514A-2 (1-520) x US-09-326-203A-1 (1-1942)

QY 1 MetalalleuaspSerAlaGlyValThrThrValThrGluAsnGlyGlyGluPhe 20
|||||


```

Db      1275 CGGAGTTTATAGGATTGGTGGAGTGGAGACGTGGAGAGTACGTGGAGAAATGTGG 1334
Qy      410 AsnMetProValHisIstYrMetValAGHISileTyrPheProCysLeuArgSerLys 429
Db      1335 AATATGCCGTCATGAATGATGATGTTGGCATATATCTGTCATGCTTACAAATGGA 1394
Qy      430 IleProLysThrLeuAlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLeu 449
Db      1395 ATACCAAAAGATAGTACGATTTTGAATCCGGTTCTGTGTGTCGATTTTTCATGAGCTG 1454
Qy      450 CysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetPheGln 469
Db      1455 TCGGTGGACGTCCTTGGCAAAATATCAAGTTTGGCGCTTCCGGGTATCATGCTTCAG 1514
Qy      470 ValProLeuValPheIleThrAsnTyrLeuGlnGluArgPhe---GlySerThrValGly 488
Db      1515 GTTCCTCTGTATAGTGTGACTAATTAATCTTGCAAGAAAGTTCAAAATCAATGATGGTGGC 1574
Qy      489 AsnMetIlePheThrPheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTyr 508
Db      1575 AATATGATGTCTGGTGTCTTCTGTCTATCTTGGTCAACATATGTTGTGTGTGTAC 1634
Qy      509 TyrHisAspLeuMetLysArgLysGlySer 518
Db      1635 TACACGACTGTATGATGAATGAAAAGCTACT 1664

RESULT 15
AF129003 2099 bp mRNA linear PLN 22-DEC-1999
LOCUS Nicotiana tabacum diacylglycerol acylCoA acyltransferase (DAGAT)
DEFINITION mRNA, complete cds.
ACCESSION AF129003
VERSION AF129003.1 GI:6625652
KEYWORDS
SOURCE
ORGANISM Nicotiana tabacum (common tobacco)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
1 (bases 1 to 2099)
Bouvier-Nave, P., Benveniste, P., Oelkers, P., Sturley, S.L. and
Schaller, H.
Expression in yeast and tobacco of plant cDNAs encoding acyl
CoA:diacylglycerol acyltransferase
Eur. J. Biochem. 267 (1), 85-96 (2000)
20069349
10601854
2 (bases 1 to 2099)
Benveniste, P.
Direct Submission
Submitted (17-FEB-1999) Plant Molecular Biology Institute,
C.N.R.S., 28 rue Goethe, Strasbourg 67083, France
FEATURES
Source
1..2099
/organism="Nicotiana tabacum"
/mol_type="mRNA"
/cultivar="xanthi SH6"
/db_xref="taxon:4097"
/tissue_type="calli derived from leaf protoplasts"
1..2099
/gene="DAGAT"
/gene="DAGAT"
208..1806
/function="catalyzes the acylation of diacylglycerol by
oleoylCoA to give triacylglycerol"
/note="similar to acyl-CoA cholesterol acyltransferase"
/codon_start=1
/product="diacylglycerol acylCoA acyltransferase"
/protein_id="AAFI9345.1"
/db_xref="GI:6625653"
/translation="MTIMELPSPVEMKTTTTSGLIENLNSDLNHSVRRRSGNGTEPA
SAINSSDANMSERDRCVSGAGLETVERSKSVGESSDVIRKEDRDNDNVANGESK

```

```

STETTTTPKRAYRASAPAHRIKESPLSDAIFKQSHAGLENCVVVLAIVNSLLIT
ENIMRYGLIRAGFEFSSKSLDMDPLMCLSLQILPLAFLVERLAQORHLERAVV
TLHITTTTAAIILYPLVILIGCDSEAFELVILMVLAVICVIMKLVASYATHRORLAKS
TKDKESDQDFSYDSFKSLAYFMAVATCYOLSLSPHPCIKRGVARTLIVETG
LMGRTIEOYINPIYONSOHPLKINILAIERLKSVPRLYVWLMEFCFPHILNITL
AFELICGDEEFTKDMNNAKTIDIEYRNMMPYHKMKNVRIYPCILNGLPKVALITLA
FLVSAVFELICIAVPCRLERKMSAFMEFVQVLVITLTELORFQSSVMVNMMEFCF
CIIQPMCLLYTHDVMNRKSSAR"
BASE COUNT 543 a 405 c 485 g 666 t
ORIGIN
Alignment Scores:
Pred. No.: 2,39e-144 Length: 2099
Score: 1842.50 Matches: 361
Percent Similarity: 76.82% Conservative: 50
Best Local Similarity: 67.48% Mismatches: 100
Query Match: 66.49% Indels: 25
DB: 8 Gaps: 7

US-09-623-514a-2 (1-520) x AF129003 (1-2099)
Qy      1 MetAlaIleLeuAspSerIleGlyValIleThrValThrGluAsnGlyGlyGlyLuphe 20
Db      217 ATGGAATTCGCCGAGAGCGTCGAAATGACGACGACGACGACATTCGGGTATTGAGAAC 276
Qy      21 ValAspLeuAsp-----ArgLeuAgaArgArgLysSerArgSerAspSerAsn 37
Db      277 CTCACACTCCGATCTTAATCATCCTCGTTCGAGAGACG-----GGCACTAAT 324
Qy      38 GlyLeuLeuLeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAla 57
Db      325 GGTTTTGAGCGCGTAGTCGCAATTAACACTTCG-----GATCGCAATATGAGCGAA 375
Qy      58 AspValArgAspArgIleAspSer-----ValValAsnAspAspAlaGln 72
Db      376 GATGAGAGATGTGTGTGCGACGCGTGTGATGGAAACGTAAGTGAAGCGAGTAA 435
Qy      73 GlyThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyGly-AspAsnAsnGlyGlyArg 92
Db      436 TCGGTGTGCGAGTCCAGCATGTATGT-AATTCGAAAGGAGGACGACAGAAATGATATGCTGC 494
Qy      92 GelyGlyGlyGlyGlyArgGlyAsnAlaAspAlaThr-----PheThrTyrAr 108
Db      495 GATGTGTGAGAGAACCAATCAACGAAACCAACGACGCCGTTAAATTTACTTACAG 554
Qy      108 GProSerValProAlaHisArgAlaArgAlaArgLysSerProLysSerAspAlaIlePhe 128
Db      555 GCGCTGGCACCACGCTCACCGCGGAATCAAGAGAGTCTCTCAGCTCCGACGCCATTCTT 614
Qy      128 eLysGlnSerHisAlaGlyLeuPheAsnLeuCysValValIleAlaIleAlaValAsnSe 148
Db      615 CAACAGAGATCAACGACGCGCTTCAATCTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 674
Qy      148 rArgLeuIleIleGluAsnLeuMetLysTyrGlyTyrPheIleArgThrAspPheTrp 168
Db      675 CAGGCTGATATTCGAGAACTGATGAAATGAGTCCCTTTAAATTAAGGCGTGTGTGTGT 734
Qy      168 eSerSerArgSerLeuArgAspTrpProLeuPheMetCysIleSerLeuSerIlePhe 188
Db      735 TAGCTGAGAGTGTGAGGATGGAGTGGCTTCAATGTCTGTCTCAGCTCCAAATTTT 794
Qy      188 eProLeuAlaAlaPheThrValGlyLysLeuValLeuGlnLysTyrIleSerGluProVa 208
Db      795 GCGCGTCCGCTGCTTTCTGTGTGAGAAAGTGGACAGACGACGACATTTCACGACGCTGC 854
Qy      208 lValIlePheLeuHisIleIleIleThrMetThrGluValLeuTyrProValTyrValTh 228
Db      855 GGTGTACTCTTCACATTAATCAACACAGCTGCATTGTGTGTGTGTGTGTGTGTGTGTGT 914
Qy      228 rLeuArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeuThrCysIleVa 248
Db      915 TCTTGGGTGTGATTCCTCTTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 974

```

RESULT 14
AF298815
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

AF298815 1964 bp mRNA linear PLN 16-OCT-2000
Perilla frutescens diacylglycerol acyltransferase (Dgat1) mRNA,
complete cds.
AF298815
AF298815.1 GI:10803052
Perilla frutescens
Perilla frutescens
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; Lamiales; Lamiales; Nepentaceae; Elsholtziaceae;
Perilla.
1 (bases 1 to 1964)
Hwang, S.-K. and Hwang, Y.-S.
Isolation of Perilla frutescens diacylglycerol acyltransferase cDNA
unpublished
2 (bases 1 to 1964)
Hwang, S.-K. and Hwang, Y.-S.
Submitted (23-AUG-2000) Division of Biochemistry, NIASR, RDA, 249
Seodun-dong, Gwonsan-gu, Suwon 441-707, Korea
Location/Qualifiers
1..1964
/organism="Perilla frutescens"
/mol_type="mRNA"
/cultivar="Okdong"
/db_xref="taxon:48386"
1..1964
/gene="Dgat1"
69..1673
/gene="Dgat1"
/note="Dgat1"
/product="diacylglycerol acyltransferase"
/protein_id="AA233696.1"
/db_xref="GI:10803053"
/translation="MALDSPELLDTSSADNGAAHHTLRROARSVPPLDSDS
NSLEASAINDSERNANDLLENLGGAVESENEKQESYKEGAKVKEGTSNEN
GTDVAVKFTFRPAAPAHNRKESPLSDAIFKOSHAGLENCIIVLVANSLLIEN
LMKYGLIKSGFESSTSLRDPMLMCCLSLPVALASPLEVELVNTYLPENVAFL
HHTITVELLFPYVVLIRKDSANLSCVTTLMELFACITWKLIVSTAHNTYDIRAKSLD
KWEASRYMWDIADYVSEKSLAVFMAVPLCYOPSPRTACIRKGMVAVOLIKLVIF
TGLMGRITIDQINPIVONSOPLKMLIAIEVRLIYVNLVYVLCMEFCFPHLMTN
IALELLCFGDFREYKDMNNAVVEEYKRMNMPVHKMAYHICPCIONGIPRIYAVL
IAFLVSAIFELCVAVPCOIFKEMAFSGIMLVPIVTVTYLDEKFKNSVMGMEMFC
FCIRGPOPCVLLIYHDMRKASAR"

BASE COUNT 522 a 395 c 461 g 586 t
ORIGIN

Alignment Scores:
Pred. No.: 2,39e-145 Length: 1964
Score: 1854.00 Matches: 359
Percent Similarity: 74.18% Conservative: 49
Best Local Similarity: 65.27% Mismatches: 92
Query Match: 66.91% Indels: 50
DB: 8 Gaps: 9

US-09-623-514a-2 (1-520) x AF298815 (1-1964)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrThrValThr-----GluAsnGly 16
DB 69 ATGGCGATCTTGGACCTCGCGAGATCCGACACGACGCGTCACATGCGCAACAGGC 128
QY 17 Gt 36
DB 129 GCCGACATCACCACTCTT-----CGCCGAGAGAAAGTGCAGCGCTCCGTCGCCCT 182
QY 37 AsnGtLeuLeuLeuSerGtSerAspAsnSerProSerAspValGtAlaPro 56
DB 183 -----CTTTCGACCTCGATCCCACTCTCTG----- 209

QY 57 AlaAspValArgAspArgIleAspSerValAlaAsnAsp-----AspAlaGlnGlyThr 74
DB 210 -----GAGCAGAGAGCGCAATCAATGATTCGAAATGTTCCAAACGAC 254
QY 75 AlaAsnLeuAlaGlyAspAsnAsnGlyGly----- 84
DB 255 GCTAATTGATTCGAAATCTCCGCGGCGAGCGCGTGAATCCGAGACGAAACAGAG 314
QY 85 -----GlyAspAsnAsnGly-----GlyGtLeuGtGtGtGtGtGt 96
DB 315 AGTTAGTGTAG 374
QY 97 GtArgGt 116
DB 375 GGAACCTGATGTATGCGCGTCAATTCACATTCACAGCGCGCGCGCTGCTCAGCGG 434
QY 117 AlaArgGtLeuSerProLeuSerSerAspAlaIlePheLeuGlnSerHisAlaGlyLeuPhe 136
DB 435 AATAGAGAGAGCTCTTACCTCCAGCGCATCTTCAACAGAGCATGACGAGCTCTTC 494
QY 137 AsnLeuCyValValValIleAlaValAsnSerArgLeuIleGlnAsnLeuMet 156
DB 495 AACCTTGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 554
QY 157 LysTyrGtTyrPheLeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrp 176
DB 555 AAGTATGGGTGGCTGATCAATCAGATGATGTTGTTAGTTAGTTCAACATCGCTTAGGAT 614
QY 177 ProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaAlaPheTrpValGln 196
DB 615 CCAGCTGTAATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 674
QY 197 LysLeuValLeuGlnLysTyrIleSerGlnProValIlePheLeuHisIleIle 216
DB 675 AAGTGTGTAATCAATATATATACAGATGAGTGGTGCATCTTCTCATGTTACATC 734
QY 217 ThrMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeu 236
DB 735 ACAAGATGGAATCTGTTTCCAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 794
QY 237 SerGtValThrLeuMetLeuLeuThrCysIleValIlePheLeuLysLeuValSerTyrAla 256
DB 795 TCAAGGTCAAGCTAATGCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 854
QY 257 HisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspLys----- 271
DB 855 CAAACAACATGATGTTGAGAGACTGCAAAATCACTTATTAAGGAGAGAGAGAGAGAG 914
QY 272 -----AlaAsnProGluValSerTyrTyrValSerLeuLysSerLeuAlaTyrPheMet 289
DB 915 AGGTACTGGAACCTCGACCTACGCTTATGATGATGATGATGATGATGATGATGATGAT 974
QY 290 ValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArgGly 309
DB 975 GTTGTCTTCATGATGTTGTTACAGCAAGCTACCTCGCAGACCTGCTGCTGCTGCTGCT 1034
QY 310 TrpValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleIle 329
DB 1035 TGGGTGTAGGCACTAATTAAGCTGTAATTCACAGAGCATCATGATTTATTTATA 1094
QY 330 GluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGtLysPheLeu 349
DB 1095 GAACAGTACATTAACCGATCGTTCAAAATTTCAACATCTCTGAAAGAAACCTTTTA 1154
QY 350 TyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrValTyrPheCysMet 369
DB 1155 TATGCAATTCAGAGGCTTGAAGCTTTCGTTCAAAATTTATATATGTTGCTCTCAGC 1214
QY 370 PheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLeuCysPheGtLysP 389
DB 1215 TTTTATGTTTTCACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1274
QY 390 ArgGluPheTyrLysAspTrpTrpAsnAlaLysSerValGtLysAspTyrTrpArgMetTrp 409

AUTHORS Mielkiewska, E., Pedersen, K., Katavic, V. and Taylor, D.C.
 TITLE Direct Submission
 JOURNAL Submitted (12-MAR-2002) Seed Oil Biotechnology, National Research
 Council of Canada, Plant Biotechnology Institute, 110 Gymnasium
 Place, Saskatoon, SK S7N 0W9, Canada

FEATURES
 source
 1.2090
 /organism="Tropaeolum majus"
 /mol_type="mRNA"
 /db_xref="taxon:4020"
 171.1727
 /codon_start=1
 /product="putative diacylglycerol acyltransferase"
 /protein_id="AA03340.1"
 /db_xref="GI:20086324"
 /translation="MAVASSQNTTMSGSDSLNNERRRSPSSVIEPSSGFTST
 NGVATGHAENRQDQVAGMENATGVSLLGGGVVNGENEGVGEIDIEPTSP
 PAHRRVRESPLDAIFKQSHAGLFLNCTIVLIANSLLIENLMKYMLDITGEWF
 NSRLGDSIFKCCITLPEFLAFLIVKLVORNNHISELVAVLHYVSTAVALYPI
 VLTGDSVYMSGVMLFECIMWLKLVSTAHSSDI RTLAKSGADAHNPSTIVCS
 YDVSLSLAYFNVAPTLCYQPSYPRSSCIKGMVVOFKLVIFGLMGFIIEQYINP
 IYRNSKHPLKGDFLYIERVLYLSPNLYVMCLMPYSEFHLMLNIAELRFGDEHY
 KDMNAKTYAEYWKMMNPVHRMVRHLTFEPLRNGIPREGAIIITAFIVSGAFHELCI
 AVCHVFILMAPIGIMFOVPLVLTINYLQEKRSNSMGIMTFICILIGPMCVLLY
 YHDLINKER"

CDS

BASE COUNT 527 a 397 c 458 g 708 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 1 74e-147 Length: 2090
 Score: 1880.00 Matches: 370
 Percent Similarity: 78.11% Conservative: 44
 Best Local Similarity: 69.81% Mismatches: 88
 Query Match: 67.85% Indels: 28
 DB: 8 Gaps: 10

US-09-623-514A-2 (1-520) x AY084052 (1-2090)

QY 1 MetAlaIleLeuAspSerIleAglyValThrValThrGluAsnGlyGlyGluPhe 20
 Db 171 ATGGGGGTGGCAGAGTCGCACAGAAACAGACACA---ACCATGAGTGTACAGCGGCGAC--- 224
 QY 21 ValAspLeuAspArgLeuArgArgArgGlySerArgSerArgPhe-----SerSer 36
 Db 225 TCGGATCTTCACAAATTTCCGTAGAAAGAAACGAGATCTCCGTGATTGACCTTCGTG 284
 QY 37 AsnGlyLeuLeuLeuSerGlySerAspAsnAsnSerProSerAspArgValGlyAlaPro 56
 Db 285 TCCGGTTTACA-----TCCACCAATGGCGTACCGGCG-----ACTGCCACGCG 329
 QY 57 AlaAspValArgAspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsn 76
 Db 330 GGTGAGAAATCGTGACAGATCGGGTATGAGGAGAGAGACAGCAACAGATCGGTACAC 389
 QY 77 LeuAlaGlyAspAsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlu 96
 Db 390 TTAATTTGGA-----AATGGGAGGCGGTGTTATCGGGGAATGAA 428
 QY 97 GlyArg-----GlyAsnAlaAspAlaThrPheThrValArgProSerValProAlaHis 114
 Db 429 GAGAAACAGTAGGGGAGACATATACATTCACCTTACCGGCTTCGTTCCGGCTCAT 488
 QY 115 ArgArgAlaArgGlnSerProLeuSerSerAspAlaIlePheGlyGlnSerHisAlaGly 134
 Db 489 CGGAGGGTGAAGGAGAGCTCTTGTAGCTCTGATGCAATCTTCAACAGAGCATGGGGT 548
 QY 135 LeuPheAsnLeuGlyValValValLeuIleAlaValAsnSerArgLeuIleIleGluAsn 154
 Db 549 TTATTCACCTTGATAGTACTGCTCATTCAGTAAACAGTAGGCTTATCATTCGAAAT 608
 QY 155 LeuMetValArgIleTyrLeuIleArgThrAspPheTyrPheSerSerArgSerLeuArg 174
 Db 609 CTTATGAACTATGTTGTTGATCGATCTGTTGTTGTTAACTCAAGATCATCGGCT 668

QY 175 AspTyrProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaIlePheThr 194
 Db 669 GATGGTGCATCTTATATGCTGTCTTACACTCCCAATTTCCACTTCTGCTTTAT 728
 QY 195 ValGlyLeuLeuValLeuGlnIleGlySerGlnProValIlePheLeuHisIle 214
 Db 729 GTTGAAGACGTGTGAGGAGCAATATATGTGAACCTGTGCTTCTCTCATGTA 788
 QY 215 IleIlePheMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSerAla 234
 Db 789 ATCGTTTCCACCGTCGACATTTTATATCCAGTATGTGATCTTAAACGTGTATTCGG 848
 QY 235 PheLeuSerGlyValThrLeuMetLeuLeuThrCysIleValTyrLeuLeuValSer 254
 Db 849 TATAGTCTGAGTGTGATGATGATGCTTGTGTTGATATATGTTGTAAGCTGTGCTCA 908
 QY 255 TyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspLys-----Ala 272
 Db 909 TATGCACATCATGTTCTGATATATGAAACACAGCGCAAACTGCGCTATTAAGGGGATGCG 968
 QY 273 AsnPro-----GluValSerTyrTyrValSerLeuLeuSerLeuAlaTyr 287
 Db 969 CACCCCAATTCACCACTTGTGAGTCTCATATGATGTTAGCTTGAAGCTTTGGCATAC 1028
 QY 288 PheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArg 307
 Db 1029 TTCATGTGGCGCGCAGCATATGATATGATACAGCGTAGTATCTCGTCCGTATACCGC 1088
 QY 308 LysGlyTyrValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPhe 327
 Db 1089 AAGGGTGGGTGTTCGTCAATTTGCAACCAATTAAGTTTATAGACTCATAGGGGCTTC 1148
 QY 328 IleIleGlnGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAsp 347
 Db 1149 ATTATGACATATATTTATCTTATCTTCTGCAATTTCCAAACCCCATTTGAAGAGAT 1208
 QY 348 LeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrValTyrLeu 367
 Db 1209 TTTTATATGACATAGAAAGAGTTTGAAGCTTTCAGTTCCAAATCTATATGTTGGCTT 1268
 QY 368 CysMetPheTyrCysPhePheHisLeuTyrPheAsnIleLeuAlaGlyLeuLeuCysPhe 387
 Db 1269 TGCATGTTTACTCTTTTTCACCTCTGTTGAACATCTGCGTACGCTTCTTCGCTTT 1328
 QY 388 GlyAspArgGlnPheTyrLysAspTyrPheAsnAlaLysSerValGlyAspTyrTyrArg 407
 Db 1329 GGTGATCTGATTTCTACAAAGATTGTGGAATGCAAAACCTGTGCGAGATTTGGAAA 1388
 QY 408 MetTyrAsnMetProValHisLysTyrMetValArgHisIleTyrPheProCysLeuArg 427
 Db 1389 ATGTGGAATATGCTGCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 1448
 QY 428 SerLysIleProLysThrLeuAlaIleIleIleAlaPheLeuValSerAlaValPheHis 447
 Db 1449 AATGGATGCCCAAGAGAGGTGCATTTATTCGCTTGTGTTGCTGCTTCAT 1508
 QY 448 GluLeuCysIleAlaValProCysArgLeuPheLysLeuTyrAlaPheLeuGlyIleMet 467
 Db 1509 GAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1568
 QY 468 PheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPhe---GlySerThr 486
 Db 1569 TTTCAAGTTCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1628
 QY 487 ValGlyAsnMetIlePheTyrPheIlePheCysIlePheGlyGlnProMetCysValLeu 506
 Db 1629 GTGGGCAATATGATCTTCTGTTCACTTCTGATCACTTGGCCAACTATGTTGTTCTT 1688
 QY 507 LeuTyrTyrHisAspLeuMetAsnArgLys 516
 Db 1689 CTTATATACATGACCTGATTAATCTAAG 1718

Db 91958 ACAGGGTATTCCTGCTTTTATGAGGGTGCATCTTGATGCCACCTGATGCTGTG 92017
 QY 249 rpleuylsleuValSerTyraAlaHisThrSerTyraSpleileatSerleuAlaAsnAla 269
 Db 92018 GGGTAAGGTGGTTCTTATGCTCATACATACATACATACATACATACATACATAC 92077
 QY 269 laasplysAla 272
 Db 92078 CTGATTAAGGT-AAAATACGAAAAAGAGCGTATGTAGTCACTTGACCTGTGTTACTG 92136
 QY 273 273 AsnProGluValSerTyraValSerL 282
 Db 92137 TTTTAACCAACACTGTATGAACCTTAGGCCAATCTCGAAGCTCTCTACTAGTACGT 92196
 QY 282 eulysSerleuAlaTyraPheMetValAlaProThrleucysTyra 296
 Db 92197 TGAAGAGCTGGCATATTCATGCTGCTGCCACATGTGTATTCAGTACGTCAAAAGT 92256
 QY 296 296
 Db 92257 GCATCAACCATTCCTATACCTGCAAGATTCCTGTCTAAACCTCGATCTTCTTTTC 92316
 QY 297 297 GluProSerTyraProArgSerAlaCysIleArgIysGlyTTPValAlaArgGlnPheA 316
 Db 92317 CCAGCCAAAGTTATCCACGCTTCGATGATACGGAAGGGTGGGCTCGTCAATTTG 92376
 QY 316 316 lalysleuValIlePheThrGlyPheMetGlyPheIleIleGlu 330
 Db 92377 CAANAATGGTCATATTCACCGGATTCATGGATTTATATAGAAACAAGTACGTTTCACA 92436
 QY 330 330
 Db 92437 TCTTGCTTATTAGTTTCTCCGTGGTGAATAATCATCCCTGCTGTACACACTGACT 92496
 QY 331 331 GluTyraIleAsnProIleValArgAsnSerIysHis 342
 Db 92497 TCATGCTCTTTGTTACATTTGGCAGATATTAATCTATGTCTCAGGAATCTAAAGCAT 92556
 QY 343 343 ProIleuysGlyAspleuTyraIleGluArgValIleuylsleuSerValProAsn 362
 Db 92557 CCTTGAAGGCGATTTCTATATGCTATTTGAAGAAGTGTGAAGCTTCAGTTCCAAT 92616
 QY 363 363 LeuTyraValIlePheCysMetPheTyraCysPhePheHisLeuTrp 377
 Db 92617 TTATATGTGTGGCTGCATGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 92676
 QY 377 377
 Db 92677 TCTCTTTCAAAATATATTTGCAAAATTCGAAAAACCGAAAAAGCTTAATCTCATACGAAT 92736
 QY 377 377
 Db 92737 TGAATATTTTACTTCTAGAGTCGGTATGTAATTTGAGTACGTAACGCAAAATCTCT 92796
 QY 378 378 LeuAsnIleleuAlaGluLeuLeuCysPheGlyAspArgIleuPheTyra 394
 Db 92797 GTCCAAAGGCTTAACATATTTGGCAGAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCA 92856
 QY 395 395 AspTTPTPAsnAlaIleuSerValGlyAsp 404
 Db 92857 GATTGTGGAAATGCAAAAAGTGGGAGA-TGTGAGCTATTTTACTCAAAAAGAAACTTA 92915
 QY 404 404
 Db 92916 TGAATTTTATATGTTGTGTTTGGGTCATCTAATCAACCAATTCATGATTCACT 92975
 QY 405 405 TyrtTPArgMetTTPAsnMet 411
 Db 92976 GTCTTCCTTTATACAGTACGTGGAAGATGTAATATGTATGTCTCTCTCTCAACATCA 93035
 QY 411 411

Db 93036 CCTTCTTTGTACACAAAATAGAAAGAGAGCTAATTAAGATCTTGTTCCTTGACAG 93095
 QY 412 412 ProValHisIleTyraPheMetValAlaArgHisIleTyraPheProCysleuArgSerIlePro 431
 Db 93096 CCTGTTCATTAATGATGTGTGTCACATATATATCTTCCCTGCTGTGGCAGCAAGATACCA 93155
 QY 432 432 Lys 432
 Db 93156 AA-CGTGAGTACATATATACCGATATGCAATTTGTCAGATTTGTTCTGTATATAAT 93214
 QY 433 433 ThrIleAlaIleIleIleAlaPheleuValSer 443
 Db 93215 TTAACCTCCACACACTGTTTTCAGACACTGCCATATATATCTTCTCTCTCTCTCT 93274
 QY 444 444 AlaValPheHisGlu 448
 Db 93275 GCAGTCTTCATGAGGTATATACATCTTCTACATGCGCTGTCTAGACGATGAAC 93334
 QY 448 448
 Db 93335 ACCTAGTGAAGAAATGCTAATATTCAAAGCATGTGTTTACTTAACGATCTTGTGTTA 93394
 QY 449 449 LeuCysIleAlaValProCysArgleuPheIysleuTTPala 462
 Db 93395 CAATTTCTCTTTGACAGCATATGCAATGCAATGCTGCTGTCTCTCTCTCTCTCTCTCT 93454
 QY 463 463 PheLeuGlyIleMetPheGln 469
 Db 93455 TTTCTTGGGATTAATGTTTTCAGTGTAAATAATTAATTAATTAATTAATTAATTAATTA 93514
 QY 470 470 ValProleuValPheIle 475
 Db 93515 AACTTAATCTCATATCTGACCAACCAATTTGTTGAGTACGCTTGTCTCTCTCTCTCTCT 93574
 QY 476 476 ThrAsnTyraIleGlnIleuArgPheIysSerThrVal 487
 Db 93575 ACNAATATCTACAGAAAGTGTGTGCTCAACGCT-ATGCTCTCAAAACCCGAGAAATA 93633
 QY 487 487
 Db 93634 GAAGCAATTAACCTTCTTTCATACGCTAGCCATTTAATTCGAATGCTGAACCTTAATA 93693
 QY 488 488 GlyAsnMetIlePheTrp 493
 Db 93694 ATAAAGTATCTGTTTGGAAATGGGATCATATATTTAGGTGGGAAACATGATCTCTGG 93753
 QY 494 494 PheIlePheCysIlePheGlyGlnPrometCysValIleuLeuTyraTyraHisAspleuMet 513
 Db 93754 TTCACTTCTGCAATTTTGGACAAACCATGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCT 93813
 QY 514 514 AsnArgIysGlySerMetSer 520
 Db 93814 AACCGAAAAAGATGATGTCA 93834
 QY 93814 93814
 RESULT 13
 AY084052 2090 bp mRNA linear, PLN 08-APR-2002
 LOCUS AY084052
 DEFINITION Tropaeolum majus putative diacylglycerol acyltransferase mRNA,
 complete cds.
 ACCESSION AY084052
 VERSION AY084052.1
 KEYWORDS GI:20086323
 SOURCE Tropaeolum majus (nasturtium)
 ORGANISM Tropaeolum majus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosids; eustosids II; Brassicales; Tropaeolaceae; Tropaeolum.
 REFERENCES
 1 (bases 1 to 2090).
 Mielkiewska,E., Pedersen,K., Katayvc,V. and Taylor,D.C.
 Characterization of a putative diacylglycerol acyltransferase mRNA
 from Tropaeolum majus embryo
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2090)


```

ASSIDERRRAAANRRRRRLMDPHVVKHENSPPFIIDNPLSNPNSTWCOFRNAE
LEKTDLDLRLPEHMCITPOTPRYQGLRKLILWLCKLHEHYRCQGMELAPLY
LVHVDIMRLSEVRKSYEDYFTDRDPSLMEFRDIYTFDEKQSDVONGSGSHK
NFNSIDLEDEVOVSLVMTDYSYSESELGYLSEKFMEDHAYFEDALMSGIDCFAM
ASFESYSPASGSHGLTPVLEACSAFRILAVDVSLSHSLVELGVEPOYGLMFLRY
LFGSEFLIDOLLYWDEILILADNSARFDEDSRONRPFIPDCRGTYLGMVSMILYL
RSSILSTENATCCOLRLNFPENDILAKITOKAKLIDALVLDITDMLASLNGVFDOS
NFVPARKSCSTSPRPLIIPESYWEKRMVLAKEEENKICLERQOTPEYOKKRM
LNVKLERAVIDLSHNRKIGERKANSPPVQSLLEDSQDLNCHVTAKENIHQ
ETENIMEFHSADEESYVSGSPSEESFYSLDPTSVRCSTKIENDSVSAGSNLIP
DEDEKSYISISEENSYSVSPISPAIDSNKYLCDCCTGSENNDOOTVNSPLYS
PHRNEKVPVQSDSDVSDKSVGTTKEYKILHGVQVPRKIKRLISEESHSRASA
TKTDVYIKKNOICYSSESHQALSSDSSONLRTKLNQOSMLKHEIATVYQOE
PCLYPAGLIGNLAKTLNLEKQVATITALEKRLSNLSEH*
17068..18550
/gene="At2g19250"
/note="synonym: F27F23.5"
join(<17068..17148,18012..18127,18236..18341,
18392..>18550)
/gene="At2g19250"
join(17068..17148,18012..18127,18236..18341,18392..18550)
/gene="At2g19250"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC16455.1"
/db_xref="GI:3135255"
/translation="MNDENNMFCCVCRDSRREPIDEEQSYRCLILNIVERGIS
FRNPKRLVAIYQPEGAIDQEDPVAAGIAKOLQEHCEVPRMVEKTESNHCNV
KLVSILYKCYSMTMWISLKMTELMVIMILKRKREIALAVAMEKLAAMKAT"
complement(17459..17479)
/rpf_family="(TAAAAA)n"
complement(20457..20499)
/rpf_family="AT-rich"
20539..20576
/rpf_family="(GA)n"
21713..24269
/gene="At2g19260"
/note="synonym: F27F23.6; predicted by genscan"
join(<21713..23329,23377..23431,23522..23625,23778..23868,
23909..23941,24091..24215,24233..>24269)
/gene="At2g19260"
join(21713..23329,23377..23431,23522..23625,23778..23868,
23909..23941,24091..24215,24233..24269)
/gene="At2g19260"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAC16456.1"
/db_xref="GI:3135256"
/translation="MIVESSRSSSEVENTSLNVSRYQDPFAVDTSIAESQMIHKYP
EQSVSFLKVPYTDIMNLSKPKHDESGSEKSVMKVSGKALTSILVTRRRRTLI
PGTEENILGKIRKODSLDSDIVSLYNPESTKRKNRKNCLVSRKRGSESSCTE
TGETTIRGIDGLDAFVSEHDCGERRRGRNDDCLVSRKGRVAFNSCSPSVTGR
TKISYDQADSSACOMGOIYKADSLTRPKGEIKSGHOLVYOROKOLVSNSTFE
SHVGTKRNDGRSEVLITYSRKSKSGKISGRVNGFLYTKIKFGPPARILCTE
KKNQDQSYVSRELLDDPYQVETVSCDQNDSCSLKSSSVNSCKRREDLCS
SDQSVETDIDGSSSPRQCKHCDKPGTVERKMLCDCEBEYHTRCGGVQDAEDID
EWLCPCLKNQSSKTKRGRISHKRWVVPVIGIRIGMFQADVPDMSGPMSVT
SFVEPLEIGQSEYMDLKKRKNKQCSNMVLCREEDTNGVIGCKWRRAPRSEVO

```

Alignment Scores:

```

Pred. No.: 4.22e-174 Length: 97495
Score: 2225.00 Matches: 515
Percent Similarity: 51.04% Conservative: 0
Best Local Similarity: 51.04% Mismatches: 5
Query Match: 80.30% Indels: 493
DB: 8 Gaps: 15

```

US-09-623-514a-2 (1-520) x AC003058 (1-97495)

```

Oy 1 MetalalleuaspSerAlaGlyValThrThrValThrgluAsnGlyGlyGluPhe 20
Db 90818 ATGCGATTGATTTGCTCTCGCGCTTACGAGTACGAGGAAACGGTGGCGAGAGTTTC 90877

```

```

Oy 21 ValaspleuaspArgleuArgArgArglySerArgSeraspSerAsnGlyLeuLeu 40
Db 90878 GTTCATCTTGTAGGCTTCGTCGACGGAATCGAGATCGATCTTCTTAACGACTCTT 90937
Oy 41 LeuSerGlySeraspAsnAsnSerProSeraspAspValGlyAlaProAlaAspValArg 60
Db 90938 CTCTGTGCTCCGATAAATAATCTCTTCCTGGATATGTGAGGCTCCCGCGAGCTTAGG 90997
Oy 61 AsparGlyleaspSerValAlaAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
Db 90998 GATCGGATGTATTCGCTGTTAAAGATAGCCTCAGGAAAGCCAAATTTGCCGGAGAT 91057
Oy 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyAlaArgGlyGlyGlyGlyGlyGlyAsn 100
Db 91058 AATAACGGTGGTGGCGATAAATACGGTGGTGAAGAGCGCGGAGAAAGAAAGAAAC 91117
Oy 101 AlaAspAlaThrPheThrThrArgProSerAlaProAlaHisArgAlaArgGluSer 120
Db 91118 GCCATGCTACGTTTACGATACGACCGGTGCTCAGCTCATCGAAGGCGAGAGAGAGT 91177
Oy 121 ProLeuSerSerAspAlaAlaPheLys----- 129
Db 91178 CCACCTTAGCTCCGACGCAATCTTCAACAGGTTTAAATCTCAGAAATTTGG 91237
Oy 130 -----GlnSerHis 132
Db 91238 TGTTCCTGCTGTTGTTTATATGAAATGAGTTGCTGATGTTGCTTGCATTCAGAGCAT 91297
Oy 133 AlaGlyLeuPheAsnLeuCyValValValLeuLeuAlaValAsnSerArgLeuIleLe 152
Db 91298 GCCGATATATCAACCTCTGTGTAGTAGTCTTATGCTGTAAACAGTAGACTATCATC 91357
Oy 153 GluAsnLeuMetLys----- 157
Db 91358 GAAATCTTATGAAGGTTTGCCTTACTGTTCTCTCTTTAGGAATGATGCTTCAA 91417
Oy 158 -----TyrGlyThrPleuIle 162
Db 91418 AATTATCAGACAGCAATACCTTGTGTCATTCATTCATGATGATGATGATGAT 91477
Oy 162 eArgThrAspPheThrPheSerSerArgSerLeuAspArgProLeuPheMet----- 180
Db 91478 CAGAACGATTTCTGTTTATGTTCAAGATCGCTGAGATGCGCGCTTTCATAGTGTG 91537
Oy 180 ----- 180
Db 91538 GTAAAGAGATGTTTTTATTTCAGCAATGTTACATGTTATACGATATATGATGAGT 91597
Oy 181 -----CysCysIleSerLeuSerIlePhe 188
Db 91598 TTAGTGATCAAGTCTCTTGTGATTCCTTCTGTTGTCAGTATATCCCTTTCGATCTT 91657
Oy 188 eProLeuAlaAlaPheThrValGluLysLeuValLeuGlnLysTyrIleSerGluPro-- 207
Db 91658 TCCTTTGGCTCCTTTTACGGTTGAGAAATTTGTTCTCAGAAATACATATCAGAACCTGT 91717
Oy 207 ----- 207
Db 91718 GAGTAATTACTATCTCCAGCCATTACTGTAATTTTATGAAGCAAGTTTGTATCATG 91777
Oy 208 -----ValValIlePheLeuHisIleIle 216
Db 91778 AAGCACTTACAGTCTGCTTTTGAATAATGCTCAAGGTTGTCATCTTCTTCATATTTATA 91837
Oy 216 IeThrMetThrGluValLeuLysProValIyValThrLeu----- 229
Db 91838 TCACCATGACAGAGTTTGTATCCAGTTTACGTCACCCTAAGGTGATGATGATGTTTCTG 91897
Oy 229 ----- 229
Db 91898 GTCTCAGTTTGTATGATCTGTTTAAAGTTTGTGTGACCCGCGATCTTGAATAATGG 91957
Oy 230 --ArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeuThrCysIleValT 249

```


OY 188 eProleuAlaIaphethValGluIysleuValIleuGlnIysTyrIleSerGluPro-- 207
 DB |||||
 DB 22015 TCCTTGGCTGCTTACGTTAGAGAAATTGGTACTCAGAAATACATATCAGAACTGCT 22074
 OY 207 ----- 207
 DB 22075 GAGTAATTACTATTCTCCAGCCATTACTGTAATTTTGAAGACAACTTGTATCANG 22134
 OY 208 -----ValValIlePheleuHisIleIleI 216
 DB 22135 AAGAACTTACAAAGTTCTGTTTGAAGAAAGCTCAAGTTGTCTCATCTTTCATATTATTA 22194
 OY 216 IethmetheIrgluValIleuTyrProValTyrValIthrLeu----- 229
 DB 22195 TCACCATGACAGAGGTTTGTATCCAGTTTACGTCCCTAAGGTGATCTGTTTTCTG 22254
 OY 229 ----- 229
 DB 22255 GTCTCAGTTTGTGATCTGTTTTAAAGTTTACTGTGTGACCCGGTGAATCTTGAATAAG 22314
 OY 230 --ArgCysAspSerAlaPheIleuSerGlyValIthrLeuMetLeuThrCysIleValT 249
 DB 22315 ACAGGTGATCTGCTTTTATGAGGTGCTACCTTGTATGCTCCTCCTGATTTGCTG 22374
 OY 249 rPleuIysleuValSerTyrAlaHisThrSerTyrAspIleArgSerIleuAlaAsnAla 269
 DB 22375 GGCTAAAGTTGGTTCTTATGCTCATACATGATGACATTAAGATCCCTAAGCCAAATGCAG 22434
 OY 269 lAspIysAla----- 272
 DB 22435 CTGATTAAGT-AAAAATACGAAAAAGAGCTATGTTTACTGCTCCTGCTGCTTACTG 22493
 OY 273 -----AsnProGluValSerTyrTyrValSerI 282
 DB 22494 TTTTAACCAACACGTGTATGAACCTTAGGCCAATCCTGAAGCTCTCTACTAGCTTACGT 22553
 OY 282 euIysSerIleuAlaTyrPheMetValAlaProThrIleuCysTyr----- 296
 DB 22554 TCNAGAGCTTGGCATATTTCATGATGCTGCCACATGTGTATCAGGTAACTGCAAGT 22613
 OY 296 ----- 296
 DB 22614 GCATCACCATTCTTATCTTGCAGAGCTTCTGTCTAAACCTCGGATCTTTCCTTTTC 22673
 OY 297 --GlnProSerTyrProArgSerAlaCysIleArgIysGlyTyrValAlaArgGlnPheA 316
 DB 22674 CCCAGCCAAAGTATCCAGCTCTGCATGTATACGGAAGGTTGGGTGGCTCGTCATTTG 22733
 OY 316 lAlysleuValIlePheThrIrgIyPheMetGlyPheIleIleGlu----- 330
 DB 22734 CAAAACTGGTCATATTCCACCGGATTCATGGATTATATAGAAACAAAGTACGTTTCACA 22793
 OY 330 ----- 330
 DB 22794 TCCTGCTTATTAGTTTCCCTGGTCAAAATCATCATCCCTGCTGTGACCACTTGACT 22853
 OY 331 -----GlnTyrIleasnProIleValAlaArgAsnSerIysHis 342
 DB 22854 TCATGTTCTTTGTACATTTTGGCAGTATATAAATCCTATTGTCTGAGAACTCAAGCAT 22913
 OY 343 ProIleuIysGlyAspIleuIleuTyrAlaIleGluArgValIleuIysleuSerValProasn 362
 DB 22914 CCTTGGAAAGGCGATCTTCTAATGCTATTGAAAGGTGTGAAAGCTTTCAGTTCCAAAT 22973
 OY 363 leuTyrValIThrIleuIysMetPheTyrCysPhePheHisIleuTyr----- 377
 DB 22974 TTATATGTGTGCTCGCATGTCTACTCTCTTCACCTTGGATGCTGTGATCCA 23033
 OY 377 ----- 377
 DB 23034 TCCTCTTCAAAATTAATTGCAAAATTCGAAAAACGAAAAAGCTAAATCTCATACGAATT 23093
 OY 377 ----- 377

DB 23094 TGATATTTTACTGTTCTAGAGTCGGTGAATGAATTTCAGTTACTGAAAGCAATCTCT 23153
 OY 378 -----LeuasnIleleuAlaIleuIleuCysPheGlyAspArgGluPheTyrIys 394
 DB 23154 GTCCAAAGGTTAAACATATTGGCAGAGCTTCTCTGCTTGGGGATGTGAATTTTACAAA 23213
 OY 395 AspTyrPAsnAlaIysSerValIysAsp----- 404
 DB 23214 GATGTGGAAATGCAAAAAGTGGGAGA-TGTGACCTATTTTACTCAAAAGAAACTTA 23272
 OY 404 ----- 404
 DB 23273 TGATTTTAAATGTGTGCTGTTTGGTCAATCTAATCAACCAAAATCATATTCATCT 23332
 OY 405 -----TyrTyrArgMetTyrAsnMet----- 411
 DB 23333 GTCTTCCTTTATCAGTACTGGAGAAATGTGAATGTGATGTCTCTTCTTAACATCA 23392
 OY 411 ----- 411
 DB 23393 CCTCTTTGTACACAAAATAGAAAGAGACTAATTAAGATCTGTTTCTCTGACAG 23452
 OY 412 ProValHisIysTyrPheValArgHisIleTyrPheProCysIleuArgSerIysIlePro 431
 DB 23453 CCTGTCAATAAATGATGTGTGCACATATATCTCCGCTTGGCAGCAAGATACCA 23512
 OY 432 Lys----- 432
 DB 23513 AA-GGTGAGTGAATATATACGATATGCAATTTGTCGAATTTGTTCTGTATTAAT 23571
 OY 433 -----ThrIleuAlaIleIleIleAlaPheIleuValSer 443
 DB 23572 TTAACCTCCACACACTGTTTTCAGACACCTGCATATATCTCTTCCAGTCTCT 23631
 OY 444 AlaValPheHisGlu----- 448
 DB 23632 GCAGCTTTTCATGAGGATATACATCTTCTACATTGCTCCGTCTCAAGCATGAACAC 23691
 OY 448 ----- 448
 DB 23692 ACGCTAGTAAAGAAATGCTAATATTCAAAGCATTTTCTTACTTAACATCTGTGTTTA 23751
 OY 449 -----LeuCysIleAlaValProCysArgIleuPheIysleuTyrAla 462
 DB 23752 CAAATTCCTTTGACAGCTATGATGCAGATTCCTGTGCTGCTTCAAGCTAATGGGCT 23811
 OY 463 PheIleuGlyIleMetPheGln----- 469
 DB 23812 TTTCTTGGATTAATGTCTCAGGTTAAAAAATTAATACTGCTGACGTGATTTTACTGA 23871
 OY 470 -----ValProIleuValPheIle 475
 DB 23872 AACCTAATCTCATATTCTGACCAACCAATTTGTTGATGATAGTGGCTTGTCTTCATC 23931
 OY 476 ThrAsnTyrIleuGlnIleuArgPheIysSerThrVal----- 487
 DB 23932 ACAAACTATCTACAGAAAGGTTGGCTCAAGGT-ATGCTCTCAAAACCCGAAATA 23990
 OY 487 ----- 487
 DB 23991 GAAGCAATACTCTTCTTTCATAGCCTAGCCATTTAAATCGCAATGCGAAACTTAATA 24050
 OY 488 -----GlyAsnMetIlePheTyr 493
 DB 24051 ATAAAGGTATCTGTTTGGATGGATGATATATTATAGCTGGGAAACATGATCTTCTCG 24110
 OY 494 PheIlePheCysIlePheGlyGlnPrometCysValIleuIleuTyrTyrHisAspIleuMet 513
 DB 24111 TTCACTTCTGCATTTTGGGACACGAGATGTGTCTCTTATATACACAGACCTGATG 24170
 OY 514 AsnArgIysGlySerMetSer 520
 |||||

repeat_region 16808..16936
/rpt_family="AT-rich"
complement(17023..17050)
/rpt_family="AT-rich"
17139..17462
/rpt_family="Kf:ATREP5|ATREP5 An AT-rich repetitive
sequence - a consensus."
repeat_region complement(17361..17388)
/rpt_family="AT-rich"
17599..17657
/rpt_family="AT-rich"
repeat_region 18967..18992
/rpt_family="(TA)n"
19827..19874
/rpt_family="(TA)n"
repeat_region 20755..20797
/rpt_family="(TAAAAA)n"
complement(20851..20873)
/rpt_family="AT-rich"
21014..24475
/gene="AT2919450"
/note="synonym: F3p11.5; supported by cDNA:
g1.15450799.gb.AY054480.1."
mRNA join(21014..22564,21649..21729,21821..21894,21997..22072,
22170..22237,22319..22442,22523..22600,22679..22780,
22881..23017,23162..23243,23348..23368,23453..23515,
23599..23646,23770..23832,23914..23964,24090..24475)
/gene="AT2919450"
CDS join(21175..21564,21649..21729,21821..21894,21997..22072,
22170..22237,22319..22442,22523..22600,22679..22780,
22881..23017,23162..23243,23348..23368,23453..23515,
23599..23646,23770..23832,23914..23964,24090..24194)
/gene="AT2919450"
/codon_start=1
/product="diacylglycerol O-acetyltransferase"
/protein_id="AA010144.2"
/db_xref="GI:20197480"
/translation="MAIDLSAGVTVTENGGEFVDLRLRRRKSQSSNGLLSGS
DNSSDDVGADPADVDRIIDSVNDACGTANLADGNNNGGNGRGSGRGVAD
AFETKPSVPAHRRAREPLSDAIFKSHAGLFLNLYVLLAVNSRLIENLMKMG
LRTDFWFSRSLRDMPLFMCISLISFPLAFYERKLYLOKYSIPYFIHITTM
TEVLYPVYTLRCDASAFISGVTMLTCTVWKIKVYATSDIRSLAANAADKANEV
STVYSIKSLATFKNVAPITCYOPSPRSACIRGWAROPAKLVIFEGWETIEDET
PIVRNSKHPKCDLLYALERVLKLSVPLNLYVLMCMFYCFHMLNLALAEVFEHC
YKDMWNAKSGDYWRMMNMPVHKMVRHIFPCLRSKIPKTLAIIIAFLVSAVFEHC
IAVPCRLFKLMDFELGIMFQVPLVFTNYLQEREGSTVGMMIFWETICIRGPMCVLLY
YHDLNMRKSGMS"
repeat_region 25114..25134
/rpt_family="AT-rich"
complement(25990..26015)
/rpt_family="AT-rich"
26221..26844
/gene="AT2919460"
/note="synonym: F3p11.6; predicted by genefinder;
supported by cDNA: g1.14334653.gb.AY034990.1."
26221..26844
/gene="AT2919460"
26320..26670
/gene="AT2919460"
/codon_start=1
/product="expressed protein"
/protein_id="AADI0145.1"
/db_xref="GI:4191776"
/translation="MEDYRSRGYGDGRTSDLOQYSAHRRSDGPDSESGMODLRSYS
TSTYDPTRIPEDONPKGRSSSSSSWGVDPLDQKKRKYVSRYRAATVEGKLGSRK
SPKMKDKCNKILN"
repeat_region complement(26742..26772)
/rpt_family="(CA)n"
complement(28356..31273)
/gene="AT2919470"
/note="synonym: F3p11.7; supported by full length cDNA:
Ceres.2312"
mRNA complement(join(28356..28814,28901..28989,29085..29148,

Alignment Scores:
Pred. No.: 3,95e-174 Length: 92822
Score: 2225.00 Matches: 515
Percent Similarity: 51.048 Conservative: 0
Best Local Similarity: 80.308 Mismatches: 5
Query Match: 80.308 Indels: 493
DB: 8 Gaps: 15
US-09-623-514a-2 (1-520) x AC005917 (1-92822)
CDS
29244..29326,29422..29548,29658..29742,29822..29885,
29979..30040,30154..30224,30310..30405,30479..30627,
30708..30777,30847..30882,30959..>31273)
/gene="AT2919470"
complement(join(28590..28814,28901..28989,29085..29148,
29244..29326,29422..29548,29658..29742,29822..29885,
29979..30040,30154..30224,30310..30405,30479..30627,
30708..30777,30847..30882,30959..>31273)
21 MetAlaIleLeuAspSerAlaGlyValThrThrValThrGluAsnGlyGlyGluPhe 20
21175 ATGGCGATTTTGGATTCTCTGCGCTTACTACGCTACGCGAAGACGTCGCGAGCTTC 21234
21 ValAspLeuAspArgLeuArgArgGlySerArgSerAspSerSerAsnGlyLeu 40
21235 GTGCGATCTTGATAGGCTTCTGTCGACGGAATGAGATCGGATCTTCTACGACTTCT 21294
41 LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60
21295 CTCTGCTTCCGATATAATCTCTCTGATGATGTTGGAGCTCCGCCACGTTAG 21354
61 AspArgIleAspSerValValAsnAspAspAlaGlyThrAlaAsnLeuAlaGlyAsp 80
21355 GATCGGATTCGATCCCTCTTAAACGATGACGTCACGGAACGCCAATTTGGCCGAGAT 21414
81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGly 100
21415 AATACGCTGCTGCGCATATAACGCTGCTGGAAGGCGGCGGAGAGAGAGAAAC 21474
101 AlaAspAlaThrPheThrTyArgProSerValProAlaHisArgAlaArgGluSer 120
21475 GCGAGTGTACGCTTACGATGACGCGTGGTCCAGTCATCGAGGCGAGAGAGAT 21534
121 ProLeuSerSerAspAlaIlePheLys----- 129
21535 CCACCTTAGCTCGACGCAATCTTCAACAGGTTTAAATCTCGAATTGG 21594
130 -----GlnSerHis 132
21595 TGTTCGCTTGTGTTTATATGGAATTGACTTGGTGTTCATTGCAAGCCAT 21654
133 AlaGlyLeuPheAsnLeuGlyValValIleuIleAlaValAsnSerArgLeuIle 152
21655 GCGGATTAATCACTGCTGTAGTAGTCTTATGCTGAACGTAAGCATCATC 21714
153 GluAsnLeuMetLys----- 157
21715 GAAATCTTATGAAGCTTTCGCTTACTGTTTCTCTTATAGAAATGAATTCCTGAA 21774
158 -----TyrGlyTyrLeuI 162
21775 AATTATCAGACGATTAACCTTGTGTTGCTATCATTCATGATGATGCTTGCAT 21834
162 eArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet----- 180
21835 CAGAACGATTTCTGTGTTAGTTCAAGATCGCTGAGATATGGCCCTTTCATGTGTG 21894
180 ----- 180
21895 GTAAGAAAGAGCTTTTATTTCCAGCAATGTACATGTAATGATGATGAT 21954
181 -----CysCysIleSerLeuSerIlePhe 188
21955 TTAGTCATCAAGTCTCTTCTTATCTCTTCTTCTTCTGACATATATCCCTTTCATCTT 22014

```

DB      1207 CATGATACCTTCCCTGCTGGCGATCAGATACCAAGATACCGCATTAATCATCTCT 1266
OY      440 PheLeuValSerAlaValAlaPheHisGluLeuGlySerIleAlaValProGlyArgLeuPheLys 459
DB      1267 TTCTAGAGCTCTGCGAGCTCTTCATGAGTTATGATGATGATGATGATGATGATGATGAT 1326
OY      460 LeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTrpLeu 479
DB      1327 CTATGGCTTCATGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1386
OY      480 GlnGluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheGlyIlePhe 499
DB      1387 CAGAGAAAGTTGGCTCCATGAGTGGGAAACATGATCTTGGTTCAGCTTTCGATTTTC 1446
OY      500 GlyAlaProMetCysValLeuLeuTrpHisAspLeuMetAsnArgGlySerMet 519
DB      1447 GGACACACCGATGTGTGGGCTCTTATTACCATGACCTGATGACCAAGATCATG 1506
OY      520 Ser 520
DB      1507 TCC 1509

RESULT 11
AC005917 92822 bp DNA linear PLN 27-FEB-2002
LOCUS      Arabidopsis thaliana chromosome 2 clone F3p11 map C1C06E08,
ACCESSION  AC005917
VERSION     AC005917.3 GI:20197478
KEYWORDS    HTG.
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM    Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Rosid; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 92822)
Lin, X., Kaul, S., Shea, T. P., Fujii, C. Y., Shen, M., VanAken, S. E.,
Barnstead, M. E., Mason, T. M., Bowman, C. L., Rounsley, C. M.,
Beilstein, M. I., Carrera, A. J., Creasy, T. H., Buell, C. R., Town, C. D.,
Niemann, M. C., Fraser, C. M., and Venter, J. C.
Unpublished
2 (bases 1 to 92822)
Lin, X.
JOURNAL
REFERENCE   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
AUTHORS     Medical Center Dr., Rockville, MD 20850, USA
TITLE        Direct Submission
JOURNAL
REFERENCE   3 (bases 1 to 92822)
AUTHORS     Town, C.D. and Kaul, S.
TITLE        Direct Submission
JOURNAL
COMMENT     Submitted (27-FEB-2002) The Institute for Genomic Research, 9712
FEATURES    Medical Center Dr, Rockville, MD 20850, USA, cdtown@igrr.org
SOURCE      On Apr 18, 2002 this sequence version replaced gi:6598497.
Location/Qualifiers
1..92822
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="2"
/map="C1C06E08"
/clone="F3p11"
1..27852
/misc_feature
/note="overlap with BAC clone F27F23
(AC003058:65644..97495)."
repeat_region
complement(2151..2171)
/rpt_family="AT-rich"
repeat_region
complement(5234..5268)
/rpt_family="AT-rich"
repeat_region
5488..5517
/rpt_family="AT-rich"
repeat_region
6440..8622
/rpt_family="AT-rich"
gene
/gene="At2g19420"

```

```

mRNA
CDS
/note="synonym: F3p11.2; predicted by genSCAN"
join(<6440..6458,6961..7071,8318..>8622)
/gene="At2g19420"
join(6440..6458,6961..7071,8318..8622)
/gene="At2g19420"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAD10141.1"
/db_xref="GI:4191772"
/translation="MHKEVPRKTYDSLENGYLAIYNGGDLRYFTVKIENDICVSI
YMTGTRKSRGKRIRINVRPKIAPARKKPKRKEDIMPEPSDEVDLISLGGDD
GYIPLPKKSLMLINLPDSFDEQLTMVSYLREVGKLMFG"
complement(7164..7241)
/rpt_family="(TA)n"
complement(9348..9369)
/rpt_family="AT-rich"
9624..9757
/rpt_family="AT-rich"
9956..12479
/gene="At2g19430"
/note="synonym: F3p11.3; contains a Trp Asp (WD 40) repeat
signature"
join(<9956..10281,10369..10445,10527..10572,10700..10760,
10849..10902,10993..11073,11165..11204,11477..11577,
11750..11860,12390..>12479)
/gene="At2g19430"
join(9956..10281,10369..10445,10527..10572,10700..10760,
10849..10902,10993..11073,11165..11204,11477..11577,
11750..11860,12390..12479)
/gene="At2g19430"
/codon_start=1
/product="putative WD-40 repeat protein"
/protein_id="AAD10142.1"
/db_xref="GI:4191773"
/translation="MYGDATWMNDEYESTLKERETRTVFTANAPPARISNPD
EYVASSDGTIAFHSLNSLVSSAFSGSKGDDVAYAEPRVRAHSGPAIDVKYGED
EDALLSCGDGRVGRKWRKREFAISDVSILHKNHLKPLELINPQHGPVGLASPMR
EINAMSVDPQSGSVFTAGDSCAYCWDVESGRIKMTFEGHSDYLTAVVRSASQILP
GSEEDGARIMDCGKTCYKVGISODKRSRLRVSSMALDGSSEMLVCGGKRLALNLP
ASECVGTPIPAHVODVFEDEKOVAVAGVGIVDVISQFSHLCITFRSSSL"
complement(12903..14545)
/gene="At2g19440"
/note="synonym: F3p11.4"
complement(join(<12903..13516,13584..14083,14223..>14545))
/gene="At2g19440"
complement(join(12903..13516,13584..14083,14223..14545))
/gene="At2g19440"
/codon_start=1
/product="putative beta-1,3-glucanase"
/protein_id="AAD10143.1"
/db_xref="GI:4191774"
/translation="MNLIAFYVGGIYIVVVDGIGVNMGTMATHKLPKKVYOMTKD
NINIKYKLPADDETMTSLSSGLEVYAIINPDOLKVMGSDRAKMDYHKVYTRYN
GCVNITPYAVGNRPPLKSYNGSFNLIFPALQIONALNENGLSSVAYATVPLNADYV
DSPSNPVPASGRRPDIIGQMTQIVPLFGNNSAPITINIPFLSGNDPFLNAYAR
FDGAKPVDDNCIAYTNVEDANFDVLVSLKAVGHGDMPIIVGEVGLPEGCKHANSGS
AYRFYNGILPLRGNGRPLRPYIEVLFDLDEDAKSIAPGEERHMGIFKFDGQP
KEPDLGSGGNKLLIGAEVNTYQPKKCMNTFAKDLTKAANIDVACFSDDCTALG
YSSCNTLDANGNSYAFNMYFOYKNDDEDAICITGATITTKNISOGCCNFPQIYA
STRASSFSSLYLVLLIAGWFLLSGVMEFY"
15295..17666
/rpt_family="Rf:ATREPS/ATREPS An AT-rich repetitive
sequence - a consensus"
15590..15728
/rpt_family="AT-rich"
15912..16149
/rpt_family="AT-rich"
complement(16223..16264)
/rpt_family="AT-rich"
16486..16577
/rpt_family="AT-rich"
complement(16668..16701)
/rpt_family="AT-rich"
repeat_region

```

DEFINITION Brassica napus putative diacylglycerol acyltransferase (DGAT1)
 mNNA, complete cds.
 ACCESSION AF164434
 VERSION AF164434.1 GI:5579407
 KEYWORDS
 SOURCE
 ORGANISM Brassica napus (rape)
 Brassica napus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE 1 (bases 1 to 1512)
 AUTHORS Nykiforuk, C.L., Laroche, A. and Weselake, R.J.
 TITLE Isolation and Characterization of a cDNA Encoding a Second Putative
 Diacylglycerol Acyltransferase from a Microspore-derived Cell
 Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No.
 AF164434). (PGR99-158)
 JOURNAL Plant Physiol. 121 (3), 1053 (1999)
 REFERENCE 2 (bases 1 to 1512)
 AUTHORS Nykiforuk, C.L., Laroche, A. and Weselake, R.J.
 TITLE Direct Submission
 JOURNAL Submitted (30-JUN-1999) Chemistry and Biochemistry, University of
 Lethbridge, 4401 University Drive, Lethbridge, Alberta T1K 3M4,
 Canada

FEATURES
 source location/Qualifiers
 1..1512
 /organism="Brassica napus"
 /mol_type="mRNA"
 /strain="cultivar Jet Neuf"
 /db_xref="taxon:3708"
 /note="microspore-derived cell suspension culture"
 1..1512
 /gene="DGAT1"
 1..1512
 /gene="DGAT1"
 /EC_number="2.3.1.20"
 /codon_start=1
 /product="putative diacylglycerol acyltransferase"
 /protein_id="A045336.1"
 /db_xref="GI:5579408"
 /translation="MAIDSGVAVPPTENGVAIDLRLHRRKSSSDSSNGLSDTSPS
 DVGAAAEERDVDSAAEEAGCTANLAGDAETRESAGDVFTYRPSVPAHRRRE
 SPLSDAIFKOSAGLLENLCYVAVLVNSREILTMKYGMLRDEPMSSSTLRMP
 LKPCSLSTVFPLAFTVERKYLAKRISEPVAILIIVITMTEVLIPVYTLCDSEAF
 LSGVTLMLTFCVIAWKLVSIAHTSYDIRLANSADVDEPISYVSLKAFVMAPT
 LCTQSPSPSPICIRKGVANOLAKLIVFLMGFIIIOYINPIVRNSKPLKODLLYA
 IERVKLISVNLIVMLCMFCEPHMLNIIAELCEGDSEFYEDMNNASVGDYRMM
 NNPVHKMVRHYVPCLRIRKIPVPAIIIAFLVSAVHELCIIVPRLFNIMAFNGIM
 FQVPIVEITNFIQERGSVMGNMIFESACIFQPMCGILYIHDLMNRGSMK"
 BASE COUNT 344 a 343 c 363 g 462 t
 ORIGIN

Alignment Scores:
 Pred. No.: 1,39e-183 Length: 1512
 Score: 2309.50 Matches: 442
 Percent Similarity: 88.10% Conservative: 17
 Best Local Similarity: 84.84% Mismatches: 43
 Query Match: 83.35% Indels: 19
 DB: 8 Gaps: 5

US-09-623-514a-2 (1-520) x AF164434 (1-1512)

QY 1 MetAlIleuAspSerAlaGlyValThrThr---ValThrgluAsnGlyGlyGlu 19
 DB 1 ATGGCGATTGTTGATTCTGGAGCGTCGCGTACCGCCGAGGAGAAAGCGC----- 51
 QY 20 PheValAspLeuAspArgLeuArgArgGlySerArgSerArgSerArgSerArg 39
 DB 52 GTCCGCGATCTCCACAGGCTCCACCGTGAATTCGAGTTCCTTCCAAAGGAGCTC 111
 QY 40 LeuLeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspVal 59
 DB 112 CTC-----TCCGACTACTTCCCGCTGGAGAGATGTTGGAGCTGGCGGCGGAA 159

QY 60 ArgAspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGly 79
 DB 160 AGGATCGGGTGTATTCCTCCGCGAGAGAGGCTCAGGAAACAGGAATTTTAGCT--- 216
 QY 80 AspAsnAsnGlyGlyGlyAspAsnAsnGlyGlyValArgGlyGlyGlyGlyArgGly 99
 DB 217 -----GGCGAGATGCCGAAACTNAGGAAATCCCGGAGAGC----- 252
 QY 100 AsnAlaAspAlaThrPheThrTyArgProSerValProAlaHisArgArgAlaArgGlu 119
 DB 253 -----GATTAAGCGTTTACGATTCACACCTCGCTTCCACTATCGAGAGAGAGAG 306
 QY 120 SerProLeuSerSerAspAlaIlePheGlyGlnSerHisAlaGlyLeuPheAsnLeuGly 139
 DB 307 AGTCCTCTCAGCTCCGAGCTGATCTCAACAAAGCATCGAGATGTTCAACCTCTGT 366
 QY 140 ValValValLeuLeuLeuAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyGly 159
 DB 367 GTAGTTGTTCTTCTTCTGTTACAGACATCATCATCAAAAACCTCATGAAGTATGTT 426
 QY 160 TrpLeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPhe 179
 DB 427 TGCTTATTCAGACTGATTTTGTGTTAGCTTACATCTTACGAGACTGCGGCTTTTC 486
 QY 180 MetCysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGlyLeuVal 199
 DB 487 ATGTGTTCTCTTCCACTTTCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 546
 QY 200 LeuGlnLysTyrlIleSerGluProValValIlePheLeuHisIleIleIleIleMetThr 219
 DB 547 CTTCAGAAATTCATATCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 606
 QY 220 GluValLeuTyArgProValTyArgValThrLeuArgCysAspSerAlaPheLeuSerGlyVal 239
 DB 607 GAGGTCTGTATTCAGCTGATCAGCTACAGCTACAGCTGATTCCTTCTTGTAGGTGTC 666
 QY 240 ThrLeuMetLeuLeuThyCysIleValTrpLeuLysLeuValSerTyAlaHisThrSer 259
 DB 667 ACCTTATGCTGCTCATCTTCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 726
 QY 260 TyrAspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyTy 279
 DB 727 TACGACATTAAGAACCTGCGCAATTCAGCTGAAGAGTGCATCTGTAACCTCTACTAT 786
 QY 280 ValSerLeuLysSerLeuAlaTyArgPheMetValAlaProThrLeuCysTyArgProSer 299
 DB 787 GTTAGCTGAGAGCTTGGGATTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 846
 QY 300 TyrProArgSerAlaCysIleArgLysGlyTyTrpValAlaArgGlnPheAlaLysLeuVal 319
 DB 847 TATCCAGCTTCTCCATGATCCGGAAGGTTGGTGGCTGCTGCTGCTGCTGCTGCTGCTG 906
 QY 320 IlePheThrGlyPheMetGlyPheIleIleGluGlnTyrlIleAsnProIleValAlaAsn 339
 DB 907 ATATTACAGCTGCTCATGAGATTTAATAGACCAATATTAATCTATTTGTTAGAAC 966
 QY 340 SerLysHisProLeuLysGlyAspLeuLeuTyAlaIleGluArgValLeuLysLeuSer 359
 DB 967 TCAAGACATCTCTGAAAGGAGACCTTATATGCTATGCTAAGAGAGCTTTGAAGCTTTCA 1026
 QY 360 ValProAsnLeuTyArgValTrpLeuCysMetPheTyArgPhePheHisLeuTrpLeuAsn 379
 DB 1027 GTTCCAAATCTATATATGTGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1086
 QY 380 IleLeuAlaGlnLeuLeuLysPheGlyAspArgGlnPheTyrlLysAspTrpPheAsnAla 399
 DB 1087 ATATTGGCAGAGCTCTCTGCTTGGGAGACCGTAATTCATCAAAATTTGGTGAATGCA 1146
 QY 400 LysSerValGlyAspTyArgMetTrpAspMetProValHisLysTrpMetValArg 419
 DB 1147 AAAAGCGTGGAGATTTATGGAGAAATGTGGAATATGCTGCTTCAAAAGATGAGTTTCCA 1206
 QY 420 HisIleTyPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIleAla 439

AUTHORS Brown, A.P., Schlerer, T.P. and Slabas, A.R.
 TITLE Direct Submission
 JOURNAL Submitted (04-Apr-2000) Biological Sciences, University of Durham,
 South Road, Durham DH1 3LE, UK
 FEATURES Location/Qualifiers

SOURCE

1. 1537
 /organism="Brassica napus"
 /mol_type="mRNA"
 /culturivar="Commonche"
 /db_xref="taxon:3708"
 /dev_stage="embryo"
 22. 1527
 /EC_number="2.3.1.20"
 /codon_start=1
 /product="putative diacylglycerol acyltransferase"
 /protein_id="AAF64065.1"
 /db_xref="GI:7576941"
 /translation="MELIDSGGVMPENGGADLDTLRRKPRSSNGLLPDSTVS
 DADYRDRDSAVEDTQGRANLAGEHEIESGEGANGVDYRTTPSPVARRVRES
 LSSDAIEKQSHAGLENLGVVAVANSRLILENLKXGWLIRTDVFWSTSLRMLPL
 MGCCLSLIFPLAFTVEKLVQKCISEPVVILHILITMTEVLPPVYVTLKDSAFIS
 GVPLMLLCIYVAKLVSAHNYDIRTLANSDDKAPSVYVSLKSLAYMLAFTLC
 YOPSPRSPCRKGVVARQPAKLIIFGFMGITIOYINPIVRSKKHPLKGDILYGE
 RVKLSVPLVYVLCMPYCPFHMLNTIAELICGDRDFEYCDMNAKSVGYRRRN
 PVHKMRAHYFPCLRRNTPKVPVAILLFLYSVAFHCLVAPCLRLKMLVFLGIMFO
 VPLVFTINYLDERGSMVGNMIFWTFCTIFGPMCVLLYHDLNMRKMS"

CDS

BASE COUNT 371 a 322 c 355 g 489 t
 ORIGIN

Alignment Scores:

Pred. No.: 4.38e-187 Length: 1537
 Score: 2351.50 Matches: 447
 Percent Similarity: 89.23% Conservative: 17
 Best Local Similarity: 85.96% Mismatches: 37
 Query Match: 84.86% Indels: 19
 Gaps: 5

US-09-623-514a-2 (1-520) x AF251794 (1-1537)

1 MetAlaIleuAspSerAlaGlyValThrThrValThrGluAsnGlyGlyGluPhe 20
 22 ATGGAGATTGGATTGCGAGGCGTCACTATGCCAGCGAGAACGGTGGT----- 72
 21 ValAspLeuAspArgLeuArgArgGlySerArgSerArgSerArgSerArgLeu 40
 73 GCCGATCTCGATACCTCTGTCACCGGAACCGAGATCGATCTTCCAAATGGACTTCT 132
 41 LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60
 133 -----CCTGATTCGTAACGTGTTCCGAT-----GCTGACCTGAGG 168
 61 AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
 169 GATCGGGTGTGATTCAGCTGTT---GAGGATCTCAAGAAAGCAATTTGGCCGAGAAA 225
 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyValArgGlyGlyGluGlyArgGlyAsn 100
 226 AACGAA-----ATTAGGGAATCCGGTGGAGAGCGGGGCGGAGAAC 264
 101 AlaAspAlaThrPheThrThrArgProSerValProAlaHisArgArgAlaArgGluSer 120
 265 GTGATGTAGCTACAGTATCGCGCGTTCAGCTCATCGAGGGTCCGCGAGAGT 324
 121 ProLeuSerSerAspAlaIlePheGlyGlnSerHisAlaGlyLeuPheAsnLeuGlyVal 140
 325 CCACCTACGCTGTGACGCCATCTTCAACAGAGCCCTGCTGACATATTAACTGTGTGA 384
 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetGlyTyrGly 160
 385 GTAGTTCTTGTGGTGAACAGTAGACTCATCAAGAAATTCATGAGAGCGTTGG 444
 161 LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180

445 TTGATCAACAGTATTGTTGTTAGTTCACAGCTCTCGGAGATTTGGCCCTTTTCATG 504
 181 CysCysIleSerLeuSerIlePheProLeuAlaIlePheThrValGlyLeuValLeu 200
 505 TGTGTCTCTCCCTTTCATATCTTTCCTTTGGCTGCTTACCGTGCAGAAATTTAGTACT 564
 201 GlnIleTyrIleSerGlnProValIlePheLeuHisIleIleIleThrMetThrGlu 220
 565 CAGAAATGCATATCGAACCTGTTGTCATCATATTTCTCATATATATATACCATGACGAG 624
 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
 625 GTCTGTATCCAGTATATGTCATCTTAAGTGTGATTCGCCCTTATACAGTGTACAG 684
 241 LeuMetLeuLeuThrCysIleValThrPheLeuValSerTyrAlaHisThrSerTyr 260
 665 TTGATGCTCCCTACCTTGATTTGTGCTGAGAGTTGTTCTTACGCTCATATCTACAT 744
 261 AspIleArgSerLeuAlaAsnAlaAlaAspIleValAsnProGluValSerTyrTyrVal 280
 745 GACATTAAGAACCCCTAGTAATTCATCTGATAAGGCCAATCCGAAAGTCTCCATAGT 804
 281 SerLeuIleSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyr 300
 805 AGCTTGAAGACCTGGCGTATTTCATGCTGTCTCCACATTTGTTATCACCCGAGCTAT 864
 301 ProArgSerAlaCysIleArgGlyGlyThrValAlaArgGlnPheAlaLysLeuValIle 320
 865 CCACCTTCCATGATATCCGGAAGGTTGGGTGCTGCTCATATTTGCAAAAGTATCTATA 924
 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValAlaArgAsnSer 340
 925 TTCATGATTCATCATGATATTATATATAGACAAATATATATATATATATAGTAAGAC 984
 341 LysHisProLeuIleGlyAspLeuLeuTyrAlaIleGluValGlyLeuLeuSerVal 360
 985 AAACATCTTTGAAAGGGGATCTTATACGCTGTGGAAGAGTGTAAACCTTTCACTT 1044
 361 ProAsnLeuTyrValThrPheLeuCysMetPheTyrCysPhePheHisLeuThrLeuAsnIle 380
 1045 CCAATTTATACGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1104
 381 LeuAlaGluLeuLeuCysPheGlyAspArgGlnPheTyrIleAspTrpTrpAsnAlaLys 400
 1105 TTGGCAGAGCTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1164
 401 SerValGlyAspTyrThrArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
 1165 AGCGTGGAGATTATTTGGAAGATGGAATATGCTGTTCAATTAATGATGCTTCGACAT 1224
 421 IleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPhe 440
 1225 GTATPCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1284
 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460
 1285 TTAGTCTGCGAGCTTTCATGAGTATGATGATGATGATGATGATGATGATGATGATGAT 1344
 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln 480
 1345 TGGGCTTCTTGGGATATATGTTTCAGGTGCTTGGTATTAATTCACAACTACTACAA 1404
 481 GluArgPheGlySerThrValGlyAsnMetIlePheThrPheIlePheCysIlePheGly 500
 1405 GAAAGTTTGGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1464
 501 GlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgGlySerMetSer 520
 1465 CAACGAGATGTGTGCTTCTTATATATACAGACCTTGAAGAACCGAAGAAAGATGTCA 1524

RESULT 10
 AF164434
 LOCUS

AF164434

1512 bp

mRNA

linear

PLN 30-NOV-1999

```

/translation="MAILDSAGVTVTEENGSEFVDLRLRRKSSDSSNGILLSSG
DNNSPSDDVGAADPADVRIDSDVNDADQAGLANAGNNGGRRGGREGAD
AFPTVSPVAHRRARSPSLSDAIFKQSHAGFLNLCVGLIVNSRLIENMKGM
LIRTFDESSRSRLSDWPLFMWCSISLIFPLAFTVERKIVLOXISEPVYELIITM
TEVLYPVYTLRCDSDAFISGVTSMILITCIYMLKIVSAHNSYDRLSANAADKANEY
SYVYSKSLAFVAVPTLCOPSPRSACIRKGVAROKKIVITFGPKFTIEQITN
PIYRNSKHPKGLDLVLAIEVKLSVFNLYVLMCTYCFEHLNMLIAELLCGDEEF
YKDMWNAKSVGDYRMMNMVHRKMWVRIHYFPCLRKIPKTLAIIIAFLVSAVHEIC
IAVPCRLFLKMLAFGLIMFOVPLVFITNYLOERGSVGNMIFWIFICIFGPMCVILLY
YHDLNMRKGSMS"
BASE COUNT      499 a      413 c      447 g      646 t
ORIGIN
Alignment Scores:
Pred. No.:      5,28e-222      Length:      2005
Score:          2771.00      Matches:      520
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             8      Gaps:      0
US-09-623-514a-2 (1-520) x AY054480 (1-2005)
QY      1 MetAlAlIeLeuSpSerAlaGlyValThrValThrGluAsnGlyGlyGluPhe 20
DB      162 ATGGCGATTTTGAATTCCTGCTGGCGTTACTACGCTGACGAGAACGGTGGCGGAGACTTC 221
QY      21 ValAspLeuAspArgLeuArgArgArgSerArgSerAspSerSerSerSerSerSer 40
DB      222 GTGCAGCTTGAATAGCGCTCTCGACGGAATCGAGATCGGATCTTCTTAACGAGACTTCT 281
QY      41 LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60
DB      282 CTCCTCGCTCCCATATATATCTCCCTCGATGATGTTGAGACTCCCGCGGACGTTAGG 341
QY      61 AspArgIleAspSerValValAsnAspAspAlaGlyIleThrAlaAsnLeuAlaGlyAsp 80
DB      342 GATCGGATTTATTCCTGTTAAAGATGACGCTGAGGAAACGCCAATTTGGCCGGAGAT 401
QY      81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGly 100
DB      402 AATAAGCGGTGGTGGCATATATACGGTGGTGGAAAGCGCGCGGAGAGAGAGAGAGAAC 461
QY      101 AlaAspAlaThrPheThrTyraGProSerValProAlaHisArgArgAlaArgGluSer 120
DB      462 GCCGATGCTACGTTCGTACGACGCTCGCTCCAGCTCATCGGAGCGGAGAGAGAGT 521
QY      121 ProLeuSerSerAspAlaIlePheGlyGlnSerHisAlaGlyLeuPheAsnLeuGlyVal 140
DB      522 CCACCTTAGCTCCGACGCAATCTTCACAGAGCCATGCCGGAATTATTCACACCTCTGTGA 581
QY      141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyGlyTrp 160
DB      582 GTAGTCTTATCTGCTGTAACAGTACGATCATCATGAAATCTTATGAAGTATGTTGG 641
QY      161 LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180
DB      642 TTATATCAGAACGCAATTTCTGTTAGTTCAGATCGCTGCAATTCGCCGCTTTTCATG 701
QY      181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200
DB      702 TGTGTATATCCCTTCGATCTTCTTCCTTGGCTGCTTTCGCTTGAAGAAATGGTACTT 761
QY      201 GlnLysTyrlIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220
DB      762 CAGAAATACATATACAGAACCTGTTCATCTTCTTCATATATATACACATGACAGAG 821
QY      221 ValLeuTyrrProValTyrrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
DB      822 GTTTTGTATCCAGTTTACGTCACCCCTAAGGCTGATTTCTCTTTTATAGGTGTCAC 881
QY      241 LeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTyrrAlaHisTrpSerTyrr 260

```

```

DB      882 TTGATGCTCTCACTTCGATTTGTGGCTAAAGTTGGTTTCTTATGCTCATACTACTAT 941
QY      261 AspIleArgSerSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrrVal 280
DB      942 GACATATAGATCCCTAGCAATTCGACCTGATTAAGCCCATCTTAAGTCTCTCTACGTT 1001
QY      281 SerLeuLysSerLeuAlaTyrrPheMetValAlaProThrLeuGlyTyrrGlnProSerTyrr 300
DB      1002 AGCTTGAAAGAGCTTGGCATTTTCATGCTGCTGCCACATGTTGTTATACGCCAAGTTAT 1061
QY      301 ProArgSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIle 320
DB      1062 CCAGTTCCTGCAATATACGAAAGGTTGGTGGCTCTCAATTTGCAAAAGTGCATAT 1121
QY      321 PheThrGlyPheMetGlyPheIleIleGlnGlnTyrrIleAsnProIleValArgAsnSer 340
DB      1122 TTCACCGGATTCATGGGATTTATATATGAACCAATATATATCTTATGTCAGAACTCA 1181
QY      341 LysHisProLeuLysGlyAspLeuLeuTyrrAlaIleGluArgValLeuLysLeuSerVal 360
DB      1182 AACCATCTTTGAAAGCGCATCTTCATATATGCTATATGAAGAGTGTGAAGCTTTCAGTT 1241
QY      361 ProAsnLeuTyrrValTrpLeuGlyCysMetPheTyrrCysPhePheHisLeuTrpLeuAsnIle 380
DB      1242 CCAAAATTTATATGTGGCTCTGCATGTTCTACTGCTTCTCCACCTTTGGTTAAACATA 1301
QY      381 LeuAlaGluLeuLeuGlyCysPheGlyAspArgGluPheTyrrLysAspTrpTrpAsnAlaLys 400
DB      1302 TTGGCAGAGCTTCTCTCTCTCTCGGGATCGCAATTCACAAAGTTGCTGCAATGCCAAA 1361
QY      401 SerValGlyAspTyrrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
DB      1362 AGTGTGGAGATATACGGAAGAAATGTGGAATATCCCTGTCATTAATGAGATGTTCCAGAT 1421
QY      421 IleTyrrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIleAlaPhe 440
DB      1422 ATATACTCCCGGTGCTGGCCAGCAAGATATCCAAAGACCTCCCATATCATTTGCTTTC 1481
QY      441 LeuValSerAlaValPheHisGluLeuGlyCysIleAlaValProCysArgLeuPheLysLeu 460
DB      1482 CTAGTCTCTGACAGCTTTCATGAGCTATGATGATGAGCTGCTGCTGCTCTCAAGCTTA 1541
QY      461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrrLeuGln 480
DB      1542 TGGGCTTTTCTTGGGATATATGTTTCAAGTCCCTTGGTCTTCAACAACTATCTACAG 1601
QY      481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500
DB      1602 GAAAGGTTTGGCTCAACGGTGGGGAACATATGCTTCTGCTTCATCTTCTGATTTTCGA 1661
QY      501 GlnProMetCysValLeuLeuTyrrTyrrHisAspLeuMetAsnArgLysGlySerMetSer 520
DB      1662 CAACCGATGTGTGGCTTCTTATATCCACAGACTGATGAAACGGAAGAGATGAGTCTCA 1721

```

RESULT 9
 AF251794
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Brassica napus (rape)
 Brassica napus
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE
 1 (bases 1 to 1537)
 BROWN, A.P., SCHLEIFER, T.P. and SLABAS, A.R.
 Characterization of a putative diacylglycerol acyltransferase cDNA
 from Brassica napus embryo
 JOURNAL
 Unpublished
 2 (bases 1 to 1537)

OY	81	LSNANSGLYGLYGLYASPSANSGLYVALGYLVRGGLYGLYGLYGLYGLYVARGLYASN	100
Db	471	AATAACGGTGCGCGCATATAACGGCTGCTGGAAGGCGCGGGAAGGAAGAAAC	530
OY	101	ALAASPAALATHRPHETHRYTARGPROSERVALPROALANISARGALAAARGLSER	120
Db	531	GCCGATGCTAGCTTTAGCTATCGAACCGTCGGTTCAGCTCATCGAAGCGGAGAGAGAGT	590
OY	121	PROLEUSERSERASPAALIEPHELYSGLINSETHSALAGLYLEUPHEASNLSCYSVAL	140
Db	591	CCACTTACTCTCGAGCGCAATCTTCAAAAGAGCCATGCCGGAATTATTCACCTCTGTCTA	650
OY	141	VALVALLEUILLEALVALASNSEARCPHEULIELEGUANLEUMELVITYGLYTRP	160
Db	651	GTAGTCTTATTCCTGCTAACAAGTAGTACTATCAATCGAAATCTTATGAAATGATAGGTGG	710
OY	161	LEULIEARGTHRASPHETTRPHESESERARGSERLEUARGASPTTRPROLEUPHEMET	180
Db	711	TTGATTCAGAACGGAATTCGTGGTTTAGTTCAAGATCGGTGCCAGATTGGCGCCTTTCTATG	770
OY	181	CYSGLYSILIESERLEUSERLIEPHEPROLEUALAALAPHEHTRVALGLULYSLEUVALLEU	200
Db	771	TGTTGTAATATCCCTTTCGATCTTTCCTTTGGCTCCCTTACGGTTGAGAAATGTGTTACTT	830
OY	201	GLINLYSTRILIESERGLUPROVALVALIEPHELEUNHSLIELEIETHMETETHGLU	220
Db	831	CAGAATATCATATACAGAACCTGTGTGTCACTCTTCTCAATATATATATACCATATACAGAG	890
OY	221	VALLEUTYTRPROVALTYRYVALTHRLEUARGCYASPSERALAPHEUSERGLYVALTHR	240
Db	891	GTTTTGTATCCAGTTTACGTACCCCTAAGGTGATTCGTCTTTTATATCGAGTGTCACT	950
OY	241	LEUMETLEULEUTHRCYSLIEVALITRPLEULYSLEUVALSERTYRALANHSTHRSERTYR	260
Db	951	TTGATATGCCCTCACCTTGATGTGTGGCTAAAGTTGGTTCTTATGCTCATACATCACTAT	1010
OY	261	ASPLIEARGSERLEUALAASNALIALASPRYSLAASNPROGLVVALSERITYRYVAL	280
Db	1011	GACATTAAGATCCCTTAGCCCAATGCAGCTGATAGGCCAATCCGTGAAGTCTCTACTACGTGT	1070
OY	281	SERLEULYSERLEUALATYRPHMETVALALAPROTHRILEUCYSTYRGLINPROSERTYR	300
Db	1071	AGCTTGAAGACCTTGGCATATTTCATGCTGCTCCACATGTGTATTACGCCAAGATTAT	1130
OY	301	PROARGSERALACYSILIEARGLYGLYTRPYVALAARGGLNPHLEALALYSLEUVALILE	320
Db	1131	CCAGTCTCTCATGTATACGGAAGGGTGGGTGCTGCTCAATTGGCAAACTGGTCATATA	1190
OY	321	PHETHRGYPHEMETGLYPHEILIEGLINGLNYRILLEANPROILIEVALARGANHSER	340
Db	1191	TTTACCCCGATTCATCGGATTTATATATAGAACAAATATATATTAATCTTATGTTCAGGAACATCA	1250
OY	341	LYSHISLPROLEULYSGLYASPLEULEUTHRYALALIEGLUARVALLEULYSLEUSERVAL	360
Db	1251	AAGCATCTCTTGAAGAAGCGATCTCTATATGCTATTTGAAGAGGTGGAACCTTTCAGTT	1310
OY	361	PROASNLEUTHRYALITRPLEUCYSMETPHEITYRCYPHEPHEHNLSEUTRPLEUASNILE	380
Db	1311	CCAAATTTATATATGTGTGGCTCTGCATGTCTACGTCTTCACGCTTTCACACCTTGGTTAAACATA	1370
OY	381	LEUALAGLULEULEUCSPHNEGILYASPARGLUBPHERYLYLSASPTPTTPASNALALYS	400
Db	1371	TTTGGCAGAGCTTCCTCTGCTCGGGGATCGTGAATTCTTACAAAGATTTGGTGAATGCAAAA	1430
OY	401	SERVALGLYASBPYRTRPARGMETTRPASNMEPROVALHSLYSTPMETVALARGHS	420
Db	1431	AGTGTGGAGATTACTGGAGATGTGGAAATATGCTCTGTCAATTAATGATGGTTCGACAT	1490
OY	421	ILEUTHRPHETPCYSLSEARGSERLYSILEPROLYSTHRILEUALALIELEIETHLAPHE	440
Db	1491	ATATACCTTCCCGTCTGCTGGCAGCAAGATATCCAAAGACACCTGCCCATTTATCATTTGCTTTC	1550
OY	441	LEUVALSERALVALPHEHISGLULEUCYSILIEALVALPROCYASARGLEUPHELYSLEU	460

Db	1551	CTAGTCTCGACGTCCTTCACTACACTATGCATCCAGTCTCCTTGCTTAACCTA	1610
Oy	461	TTPAlaphleuglyIlemetPhegINvalProleuValPheileThraAsnTyLeuGlN	480
Db	1611	TGGGCTTTCTTGGGATATATTCTTTCAGGTGCTTGGCTTCATCACACAACATATCTACAG	1670
Oy	481	GLUArgPheglySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly	500
Db	1671	GAAAGCTTTGGCTCAACGCTGGGGAACATCATCTTCCTGATCTCATCTTCGATTTTGGGA	1730
Oy	501	GIInPrometCysValIleuLeuTyThrHisaspMetIsnArgLysGlySerMetSer	520
Db	1731	CACCGATGTGTGTGCTCTTATTACACGACCTGATGAACCGAAAAAGATCGATGTCA	1790
RESULT 8			
LOCUS	AY054480	2005 bp mRNA linear PLN 05-SEP-2001	
DEFINITION	Arabidopsis thaliana diacylglycerol O-acetyltransferase [At2g19450; F3p11.5] mRNA, complete cds.		
ACCESSION	AY054480		
VERSION	AY054480.1	GI:15450799	
KEYWORDS	FLI CDNA.		
SOURCE	Arabidopsis thaliana (thale cress)		
ORGANISM	Arabidopsis thaliana		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 2005)		
AUTHORS	Southwick,A., Karlin-Neumann,G., Nguyen,M., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banb,J., Carninci,P., Chen,H., Cheuk,R., Chung,M.K., Hayashizaki,Y., Ishida,J., Kamlya,A., Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K., Savano,H., Sekurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K., Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.		
TITLE	Direct submission		
JOURNAL	Submitted (28-AUG-2001) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA		
COMMENT	e-mail for correspondence: arabsequence.stanford.edu		

RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA : "RIKEN Arabidopsis Full-length cDNA") : Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Nguyen,M., Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J., Bowser,L., Jones,T., Banb,J., Chen,H., Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P., Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

Southwick,A., (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W. (SSP/Stanford) contributed equally to this work as PIs.

Location/Qualifiers

- 1..2005
 - /organism="Arabidopsis thaliana"
 - /mol_type="mRNA"
 - /db_xref="taxon:3702"
 - /chromosome="II"
 - /clone="RAF107-14-014"
 - /note="This clone is in pBluescript"
 - 1..2005
 - /gene="At2g19450; F3p11.5"
 - 162..1724
 - /gene="At2g19450; F3p11.5"
 - /codon_start=1
 - /product="diacylglycerol O-acyltransferase"
 - /protein_id="AKR6671.1"
 - /db_xref="GI:15450800"

FEATURES

SOURCE

gene

CDS

Db 717 TTGATCAGAACGATTTTCGTTTACGTACAGATCCGCGAGATTGCGCGCTTTCATG 776
 QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLeuValLeu 200
 Db 777 TGTGTATATCCCTTCGATCTTCCCTTGGCGCTTACGGTTGAGAAATGGTACTT 836
 QY 201 GluLeuTyrIleSerGluProValIlePheLeuHisIleIleIleThrMetThrGlu 220
 Db 837 CAGAAATACATATCAGAACCTGTGTGATCTTCTTCATATATATATACACATGACAGAC 896
 QY 221 ValLeuTyrProValIleValIleThrLeuArgCysAspSerIlePheLeuSerGlyValThr 240
 Db 897 GTTTGTATACCACTTACGTCACCCATGAGGTGTGATTCCTTTTATCAGGTGCTACT 956
 QY 241 LeuMetLeuLeuThrCysIleValIleTyrLeuLeuValSerTyrAlaHisThrSerTyr 260
 Db 957 TTGATGCTCTCTACTGTCATGTGTGCTAAATGTTGTTCTTATGCTCTACTACTAT 1016
 QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspIleAsnProGluValSerTyrTyrVal 280
 Db 1017 GACATAGATCCCTACCAATGACGATGATAGGCCAATCTCGAAGTCTCTACTAGCTT 1076
 QY 281 SerLeuLeuSerLeuAlaTyrPheMetValAlaProThrLeuLeuTyrGlnProSerTyr 300
 Db 1077 AGCTTGAAGAGCTTGGCATATTTTCATGTGCTCCACATGTTATCAGCCAGTTAT 1136
 QY 301 ProArgSerAlaCysIleArgGlyTyrPheValAlaArgGlnPheAlaValLeuValIle 320
 Db 1137 CCACGTTTCGATGATAGGAGAGGTGGGTGGCTGCTCATATTGCCAAACATGGTCTATA 1196
 QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValAlaGlnSer 340
 Db 1197 TTCACCGGATTCATGGGATTATATAGAACATATATATATCCATTGTGCAGAACTCA 1256
 QY 341 LysHisProLeuLeuGlyAspLeuLeuTyrValIleGluArgValLeuLeuSerVal 360
 Db 1257 AACGATCCTTTGAAGCCGATCTTCTATATGCTATTTGAAGAGGTGTGAGCTTTCAGT 1316
 QY 361 ProAsnLeuTyrValIleTyrLeuCysMetPheTyrCysPhePheHisLeuTyrLeuAsnIle 380
 Db 1317 CCAAATTTATATGTGGCTCTGATGCTCTACTGCTTCTCCACCTTGGTAAACATA 1376
 QY 381 LeuAlaGluLeuLeuLeuGlyAspArgGluPheTyrIleAsnProIleValAlaGlnSer 400
 Db 1377 TTGGCAGACCTCTCTGCTTGGGATCGGATTCATACAAATGCTGGGATGCAAAA 1436
 QY 401 SerValGlyAspTyrTyrPheMetTyrPheMetProValHisIleTyrPheMetValArgHis 420
 Db 1437 AGGTGGGAGATTTACGAGAAATGTGCAATATGCTGCTTCATTAATGATGTTGCGCAT 1496
 QY 421 IleTyrPheProCysLeuArgSerIleProIleThrLeuAlaIleIleIleAlaPhe 440
 Db 1497 ATTTACTTCCCGGCTTGGCGACCAAGATACCAAAAGACACTGCCATTATCATGTGCTTC 1556
 QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheIleLeu 460
 Db 1557 CTAGTCTCGCACTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1616
 QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln 480
 Db 1617 TGGGCTTTTCTTGGGATTTATGTTTTCAGGTGCTTTCCTTCATCAAACTATCTACAC 1676
 QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTyrPheIlePheCysIlePheGly 500
 Db 1677 GAAAGGTTTGGCTCAACGGGTGGGAAACATGATCTTCTGTTATCTTCGCAATTTTCGGA 1736
 QY 501 GlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgIleGlySerMetSer 520
 Db 1737 CAACCGATGTGTGTCTCTTATATACCAAGACCTGATGATGATGATGATGATGATGAT 1796
 RESULT 7
 ATH131831
 LOCUS ATH131831 1988 bp mRNA linear PLN 10-JUN-1999

DEFINITION Arabidopsis thaliana mRNA for diacylglycerol O-acyltransferase.
 ACCESSION AJ131831
 VERSION AJ131831.1 GI:5050912
 KEYWORDS diacylglycerol O-acyltransferase.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE 1
 AUTHORS Hobbs, D.H., Lu, C. and Hills, M.J.
 TITLE Cloning of a cDNA encoding diacylglycerol acyltransferase from
 Arabidopsis thaliana and its functional expression
 JOURNAL FEBS Lett. 452 (3), 145-149 (1999)
 MEDLINE 99313150
 PUBMED 10386579
 REFERENCE 2 (bases 1 to 1988)
 AUTHORS Hills, M.J.
 TITLE Direct Submission
 JOURNAL Submitted (22-DEC-1998) Hills M.J., Department of Brassica and
 Oilseeds Research, John Innes Centre, Colney Lane, Norwich, NR4
 7UH, U.K.
 FEATURES
 source location/Qualifiers
 1. 1988
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /sub_species="Columbia"
 /db_xref="taxon:3702"
 231. 1793
 /EC_number="2.3.1.20"
 /codon_start=1
 /product="diacylglycerol O-acyltransferase"
 /protein_id="CA44774.1"
 /db_xref="GI:5050913"
 /translation="MAIIDSAGTYYTENGCGEPVDLRLRRKRSDDSNGLILSSG
 DNNSPDVGAPADVRDRIDSVNDAGGTANLANNNGGDDNGGKGGKGGKGNAD
 APTTTPSPVAHRRARESPSSDAIRKQSHAGFNLGVVLAIVNSRLIIEINIKYGM
 LRTDWFSSRSRLDWPLEMCISLSIFPLAATVKKLVOKYISPVVIFHIIITM
 TEVLVYVYTLRCDSAFLSGVTLMLTCTVMTLVYVAATSYDIRSLANADANAEV
 SYVSLKSLAEVAPTLCTOPSPRSACIRKGVARORAKLYIFTFMGFTIEOYIN
 PIYRNSKHPLKGLDLAIVRIKLVSPNLYWCMGCEPHMLNLALFLRGDDEF
 YKDMANAKSGVDGWRMNNPVHNNKRRHRYIFPELRKIKRTLAIIITFLVSAPFELC
 IAVPCRLKLMARFLGTMFOVPLFTITNYLQBERGSTVGNNITFWFIFCIGDPVCVLY
 YHDLNMRKMSMS"
 ORIGIN
 BASE COUNT 478 a 423 c 436 g 651 t
 Alignment Scores:
 Pred. No.: 5,22e-222 Length: 1988
 Score: 2771.00 Matches: 520
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 8 Gaps: 0
 US-09-623-514a-2 (1-520) x ATH131831 (1-1988)
 QY 1 MetAlaIleLeuAspSerAlaGlyValIleThrValIleThrGluAsnGlyGlyGluPhe 20
 Db 231 ATGGCATTTTGGATTTCTGCTGCGTACTACGGTACGAGACGCGGAGAGTTCT 290
 QY 21 ValAspLeuAspArgLeuArgArgArgGlySerArgSerAspSerSerAsnGlyLeuLeu 40
 Db 291 GTGCATCTTGATAGGCTTGTGTCAGCGAATGAGATCGGATTCCTTAACGACTTCT 350
 QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60
 Db 351 CTCCTCGTTCGATATATATCTCTCGGATGATGATGATGATGATGATGATGATGAT 410
 QY 61 AspArgIleAspSerValValAlaAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
 Db 411 GATCGATTTGATTTCCCTGTGTTACGATGACGCTCGAGGAAACACCATTTGGCGGAGAT 470

QY	141	ValValLeuLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTrp	160
Db	657	GTACTCTTATATCGCTGAACAGTAAGACTCATCTCAAAATCTTATGAAATGATAGTGTGG	716
QY	161	LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet	180
Db	717	TTGATTCAGAACCGATTCTCGCTTTAGTTCAAAGATCGCGCAGATGTGGCCCTTTCTATG	776
QY	181	CysCysIleSerLeuSerIlePheProLeuAlaIlePheThrValGluLysLeuValLeu	200
Db	777	TGTTGTATATCCCTTGGATCTTCCATTGGCTGCCCTTACGGTTGAGAAATGTGTACTT	836
QY	201	GlnLysTyrIleSerGluProValIlePheLeuHisIleIleIlePheMetGlu	220
Db	837	CAGAAATACAAATTCAGAACCTGTGTGTCATCTTCTTCATATATTATTCACCATGACAGAG	896
QY	221	ValLeuLysProValTyrValThrLeuArgLysAspSerAlaPheLeuSerGlyValThr	240
Db	897	CTTTTGATTCACGTTTACGTACCCCTAAGGCTGATTCGGCTTTTATTCAGGGTGCATCT	956
QY	241	LeuMetLeuLeuThrCysIleValTyrPheLysLeuValSerTyrAlaHisThrSerTyr	260
Db	957	TTGATGCTCCCTCATTGGCATGTGTGGCTAAAGTGGTTCTTATGCCATACATACCAT	1016
QY	261	AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTyrVal	280
Db	1017	GACATTAAGATCCCTACGCAATGACAGCTGATTAAGCCCAATCCTGAAGTCTCTACTAGCTT	1076
QY	281	SerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyr	300
Db	1077	AGCTTGAAAGACCTTGGCATATTTCATGCTGCCTCCACATTTGTATTACACCAAGTTAT	1136
QY	301	ProArgSerAlaCysIleArgLysGlyTyrPalaAlaArgGlnPheAlaLysLeuValIle	320
Db	1137	CCAGTTCCTGCATGTATATACGGAAGGTTGGGTGGCTGCTCAATTTGCCAAACCTGCTATA	1196
QY	321	PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSer	340
Db	1197	TTTACCCGGATTCATAGGATTTATATATAGAACAAATATATTAATCCTATTGTACAGAACTCA	1256
QY	341	LysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValIleLysLeuSerVal	360
Db	1257	AAGGATCTCTTGAAGGCGATCTTCTATATGCTATGTAAGAAAGATGTTGAACCTTTCAGTT	1316
QY	361	ProAsnLeuTyrValTyrPLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIle	380
Db	1317	CCAAATTTATATATGTGTGGCTGTGCATGTCTACAGCTCTTCCACCTTTGGTTAAACATA	1376
QY	381	LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLys	400
Db	1377	TTTGGCAAGAGCTTCTCTGCTCCGGGAGATCGTAAATTTCAAAAGATGGTGAATGCANAA	1436
QY	401	SerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis	420
Db	1437	AGTGTGGAGATTTCTGAGAAATGTGAATATGCTCTCTCTTAATATGAGATGGTTCGACAT	1496
QY	421	IleTyrPheProCysLeuArgSerLysIleProCysThrLeuAlaIleIleIleAlaPhe	440
Db	1497	ATATACCTTCCCGTCTCTTCCGACAGCAATACCAAGACACCTCGCATTTATCATTCGTTTTTC	1556
QY	441	LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu	460
Db	1557	CTACTCTCTGAGCTTTCATGACCTATGACATCGACATGCTTGTGCTGCTTCAAGCTA	1616
QY	461	TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln	480
Db	1617	TGGGCTTTTCTTGGGATTAATGTTTCACGTGCTTTGATCTTCATCACAAATATCTACAG	1676
QY	481	GluArgPheGlySerThrValGlyAsnMetIlePheThrPheIlePheCysIlePheGly	500
Db	1677	GAAGAGTTTGGCTCAACGGTGGGGAACAATGATCTTCTGGTTCATCTTCTGCATTTTGGGA	1736

OY		501	GlnProMetCysValIleuLeuTyrThrHisAspLeuMetAsnArgLysGlySerMetSer	520
Dd	1737	CACCCAGTCGTGGCTTTCTTTATTTACCACGACTGATGAACCGAAMAAAGATCATGTCA	1796	
RESULT 6				
AX090349				
LOCUS				
DEFINITION	Sequence 42 from Patent WO0116308.			
ACCESSION	AX090349			
VERSION	AX090349.1	GI:13444210		
KEYWORDS				
SOURCE				
ORGANISM	Arabidopsis thaliana (thale cress)			
REFERENCE	1	Lassner,M. and van Eenennaam,A. Plant sterol acyltransferases Patent: WO 0116308-A 42 08-MAR-2001; MONSANTO COMPANY (US) Location/Qualifiers 1..1942		
AUTHORS	/organism="Arabidopsis thaliana"			
TITLE	/mol_type="genomic DNA"			
JOURNAL	/db_xref="taxon:3702"			
FEATURES				
SOURCE				
BASE COUNT	461 a 421 c 425 g 635 t			
ORIGIN				
Alignment Scores:				
Pred. NO.:	5.06e-222	Length:	1942	
Score:	2771.00	Matches:	520	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	100.00%	Indels:	0	
	6	Gaps:	0	
US-09-623-514A-2 (1-520) x AX090349 (1-1942)				
OY	1	MetaIalleuenaSpSerAlaGlyValThrValThrcGuasngLygLyglUphe	20	
Dd	237	ATGGCGATTGTTGGATTCTGCTGGCGTTACTACGGTGACGGAGAACGGTGGCGGAGACTTC	296	
OY	21	ValAspleuaspArgLeuArgArgLysSerArgSerSerSerasnGlyLeuLeu	40	
Dd	297	GTCATCTTGTAGAGGCTTCGCAGCGGAAAACGAGATCGGATTCCTTTAACGGACTTCCT	356	
OY	41	LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg	60	
Dd	357	CTCTCTGGTCCGATAATTAATTCCTTCCTCGATCATGTGGAGCTCCCGCGAGCTTAGG	416	
OY	61	AspArglleasPserValValAsnAspAspAlacInglyThrAlaasnLeualGlyasp	80	
Dd	417	GATCGGATTGATTCGCCGTGTAAAGANGACCCCTCAGGGAAACGCCAATTGTGCCGAGAT	476	
OY	81	AsnAsnngLygLygLyasPasnAsnGlyGlyValArgGlyGlyGlyGlyValArgGlyAsn	100	
Dd	477	AATACGCTGCTGGCGATTAATTAACGGTGGTGAAGAGCGCGGAGAGAGAGAAAC	536	
OY	101	AlaAspAlaThrPherThrTyrArgProSerValProAlaHisArgArgAlaArgGluSer	120	
Dd	537	GCCGATCTCTACGTTACGTATCGACCGCGGCTCCACGTCATCGGAGGCGGAGAGAGCT	596	
OY	121	ProLeuSerSerAspAlaIlePheLygLnSerHisalagLyLeuPheasnLeuCysVal	140	
Dd	597	CCACTGTGCTCCGCGCAATCTTCAAACAGAGCATCCCGATTATTCACCTCTGTGA	656	
OY	141	ValValIleuIlealaValAsnSerArgLeuIleIlegLuasnLeuMetLysTyrGlyTrp	160	
Dd	657	GTAATTCTTATGCTGTGAACAGTAGACTCATCTCCAAAAATCTTTATGAAGTATGATGG	716	
OY	161	LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet	180	

OY 101 AlaaspaIatPhPheThTyraPProSerValProAlaHisArgAlaArgIuSer 120
 |||||
 DB 537 GCCGAGTCTACGTTTACGTATGACACCGGTCCAGCTCATCGAGGCGAGAGAGACT 596
 OY 121 ProLeuSerSerAspAlaIlePheIysGInserHisAlaGlyLeuPheAsnLeuGlyVal 140
 |||||
 DB 597 CCACCTAGCTCCGACCAATCTTCAACAGACGCGGATTAATCAACCTCTGTGA 656
 OY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTyr 160
 |||||
 DB 657 GTAGTCTTATGCTGTATAAGATAGACATCATCATGAAATCTTATGAAGTATGTTGG 716
 OY 161 LeuIleArgTyrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180
 |||||
 DB 717 TTGATTCAGAACGAGATTTCTGTGTAGTCAAGATCCCTCGAGATTTGCCGCTTTTCAG 776
 OY 181 CysCysIleSerLeuSerIlePheProLeuAlaIlePheThValGluLysLeuValLeu 200
 |||||
 DB 777 TGTGTATATCCCTTTCGATCTTCTTGGCTGCTTACGGGTGAGAAATTTGTAAGT 836
 OY 201 GlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleThrMetTrpGlu 220
 |||||
 DB 837 CAGAAATACATATACAAACCTGTGTCACTTCTTCATATATATATACACATGACAGAG 896
 OY 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
 |||||
 DB 897 GTTTGTATCCAGTATCTTACCTCACCCCTAAGGTGATTCGCTTTTATACAGGTGCAC 956
 OY 241 LeuMetLeuLeuThrcCysIleValIlePheLysLeuValSerTyrAlaHisTrpSerTyr 260
 |||||
 DB 957 TTGATGCTCTCACTTGCATTTGTGTGCTAAGGTGTGTTCTTATCTCATCTACTACT 1016
 OY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAspProGluValSerTyrTyrVal 280
 |||||
 DB 1017 GACATTAAGATCCCTACGCAATGAGCTGATTAAGGCCAATCCGAACTCCACTACTACT 1076
 OY 281 SerLeuLysSerLeuAlaIlePheMetValAlaProThrLeuCysTyrGlnProSerTyr 300
 |||||
 DB 1077 ACCTTGAAAGCTTGGCATATTTTCATGTGCTCCACATTTGTGTATACGCCAAGCTTAT 1136
 OY 301 ProArgSerAlaCysIleArgLysGlyTyrValAlaArgGlnPheAlaLysLeuValIle 320
 |||||
 DB 1137 CCACGCTTGCATGTATACGAGAGGTTGGGTGCTGCTGCATTTTCCAAACCTGGTCA 1196
 OY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAspProIleValArgAsnSer 340
 |||||
 DB 1197 TTCACCGGATTCATGGGATTTAATAAGAACATATAAATCCATTCACGAGACTCA 1256
 OY 341 LysHisProLeuLysGlyLysLeuLeuTyrAlaIleGluArgValLeuLysLeuSerVal 360
 |||||
 DB 1257 AAGCATCCCTTGAAGGCGATCTTCTATGCTATGAAAGAGTGTGAAGCTTTCAGTT 1316
 OY 361 ProAsnLeuTyrValTrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIle 380
 |||||
 DB 1317 CCAAAATTAATATGTGTGCTGTGATGTCTTACTGTCTTCTCCACCTTGGTTAAACATA 1376
 OY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLys 400
 |||||
 DB 1377 TTGGGAGACTTCTCTGCTTGGGGATTCGTAATTCACAAAGATTGCGAATGCAAAA 1436
 OY 401 SerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
 |||||
 DB 1437 AATGTGGGAGATTAACGAGAAATGTGCAATATGCCCTTCATAAAGATGTGTCACAT 1496
 OY 421 IleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIleAlaPhe 440
 |||||
 DB 1497 AATATCTTCCCTGCTTGGCGACAGAAATACCAAAAGACCTGCGCATTTATCATTTGCTTC 1556
 OY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460
 |||||
 DB 1557 CTAGCTCTGCAGTCTTTCATGAGCTATCATCGAGCTTCCTGTGCTCTCTCAAGCTTA 1616
 OY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln 480

DB 1617 TGGGCTTTTCTTGGGATTAATGTTTACAGTGCCCTTGTGCTTCATCAAAATCTTACAG 1676
 OY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheGlyIlePheGly 500
 |||||
 DB 1677 GAAGGTTTGGCTCAACGGGTGGGAAACATGATCTTGTGTATCTTCTGCAATTTTCGA 1736
 OY 501 GlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520
 |||||
 DB 1737 CAACGATGTGTGTCTTCTTTATTTACACAGACCTGATGACCGAAAGATCGATGTCA 1796
 RESULT 5
 AX090345 1942 bp DNA linear PAT 21-MAR-2001
 AX090345 Sequence 38 from Patent WO0116308.
 DEFINITION
 AX090345
 ACCESSION
 VERSION
 AX090345.1 GI:13444207
 KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1
 Lasserre M. and van Eenennaam A.
 AUTHORS
 Plant sterol acyltransferases
 TITLE
 Patent: WO 0116308-A 38 08-MAR-2001;
 JOURNAL
 MONSANTO COMPANY (US)
 FEATURES
 source
 1.1942 Location/Qualifiers
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 BASE COUNT 461 a 421 c 425 g 635 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5.06e-222 Length: 1942
 Score: 2771.00 Matches: 520
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 6 Gaps: 0
 US-09-623-514a-2 (1-520) x AX090345 (1-1942)
 OY 1 MetAlaIleLeuAspSerAlaGlyValIleThrValIleThrGluAsnGlyGlyGluPhe 20
 |||||
 DB 237 ATGGCATTTTGGATTCTGCTGCGCTTACTACGGTACGAGAACGTTGGCGAGAGTTTC 296
 OY 21 ValAspLeuAspArgLeuArgArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40
 |||||
 DB 297 GTTCGATCTTGATAGCTTCTGTCACGGAATGCAGATTCGATTTCTTAACGACTTCTT 356
 OY 41 LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60
 |||||
 DB 357 CTCTCGGTTCCGATTAATTAATTCCTTCGAGATGTTGGAGATCCCGCCGACGTTAG 416
 OY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
 |||||
 DB 417 GATCGGATTAATTCGTTTGAACGATGACGCTCAGAGGAACGCCAATTTGGCCCGAGAT 476
 OY 81 AsnAsnGlyGlyValAspAsnAsnGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGly 100
 |||||
 DB 477 AATTAACGATGATGCGCATTAATAACGTTGCGAAGAGCGCGGCAAGAAAGAGAAAC 536
 OY 101 AlaaspaIatPhPheThTyraPProSerValProAlaHisArgAlaArgIuSer 120
 |||||
 DB 537 GCCGATGCTACGCTTACATGACACCGTGGTTCCAGCTCATCGGAGGCGAGAGAGACT 596
 OY 121 ProLeuSerSerAspAlaIlePheIysGInserHisAlaGlyLeuPheAsnLeuGlyVal 140
 |||||
 DB 597 CCACCTAGCTCCGACCAATCTTCAACAGACGACCGGATTTATCAACCTCTGTGA 656

Db 259 CTCTGGTCCGGAATAATTCCTTCGGATGATGTTGGAGCGCCGCGACCTTGG 318
 Qy AsparglyleaspservalaValasnaspaspaalaglnclyThrAlaasnleuAlaIleasp 80
 Db 319 GATCGGATGTAATCCGCTTGTAACCATGACCTCGGGAACAGCAATTTGGCCGGAAT 378
 Qy AsnansnglyglyAspAsnansnglyglyArgglyglyIleuIleuValasn 100
 Db 379 AATAACGGTGGTGGGATTAATACGGTGGGAAGAGCGCGGAGGAAGGAAGGAAC 438
 Qy AlaaspaalathrphethrThrArgProserValProAlaHisArgAlaArgIleuSer 120
 Db 439 GCCGATGCTAGCTTACGTATCGACCGCTGGTCCAGCTCATCGAGGGCGAGAGAGT 498
 Qy ProleuserSerAspaAlaIlePheIysGlnSerHisAlaGlyLeuPheAsnLeuVal 140
 Db 499 CCACCTACCTCCGACGCAATCTTCAACAGAGCCATGCCGGAATTTACACCTCTGTGTA 558
 Qy ValValLeuIleAlaValaValasnSerArgLeuIleIleGluasnLeuMetIysTyrGlyTrp 160
 Db 559 GTAGTCTTATGCTGTAACAGTAGACTCATCATCGAATAATCTTATGAGATGTTGG 618
 Qy LeuIleArgThrAspPheThrPheSerSerArgSerIleuAlaGAspTrpProleuPheMet 180
 Db 619 TTGATCAGAACGAGATTCCTGTTAGTTCAGATCGCTGCGAGATGGCCCTTTTCATG 678
 Qy CysCysIleSerIleSerIlePhePheProleuAlaAlaPheThrValGlyIleuValLeu 200
 Db 679 TGTGTATATCCCTTCGATCTTTCCTTGGCTGGCTGACGTTGAGAAATTTGGTACTT 738
 Qy GlnIysTyrIleSerGluProValaValIlePheLeuHisIleIleIleThrMetThrGlu 220
 Db 739 CAGAAATACATATCAGAACCTGTGTGTCATCTTCTCATATATATATACCATGACAG 798
 Qy ValIleuThrProValIleThrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
 Db 799 GTTTGTATCCAGTTTACGTACCCCTAAGGTGATGTTGCTTTTATACAGGTCTACT 858
 Qy LeuMetLeuLeuThrCysIleValIleThrLeuIleValSerTyrAlaHisThrSerTyr 260
 Db 859 TTGATGCTCCCTACCTGATGCTGTGCTAAGTTGGTCTTCTTATGCTCATACATACAT 918
 Qy AspIleArgSerIleuAlaAlaAlaAlaAspIysAlaAsnProGluValSerTyrTrpVal 280
 Db 919 GACATATAGATCCCTAGCCCAATGCAAGCTGATAGGCCAATCTCGAAGTCTCTACTACGTT 978
 Qy SerIleuYsserIleuAlaIleThrPheMetValaAlaProThrLeuCysTyrGlnProSerTyr 300
 Db 979 AGCTTGAAGACCTTGACATATTTATGATGCTGCCCTCCACATGTTGTTATCACGCCAAGTTAT 1038
 Qy ProArgSerAlaCysIleArgIysGlyTrpValAlaArgGlnPheAlaIleValIle 320
 Db 1039 CCAGCTTGTGCATGATATCGAAGAGGTTGGCTGCTCTCATATTTGCCAAACTGCTCTATA 1098
 Qy PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValaArgAsnSer 340
 Db 1099 TTCACCGGATTCATGGGATTTATATAGAACATATATATATATATATGTCAGAGAACATCA 1158
 Qy LysHisProleuYsserGlyAspLeuLeuThrAlaIleGluArgValaLeuIleuSerVal 360
 Db 1159 AAGCATCTTGAAGAGCGATCTTATATGCTATGTAAGAGAGTGTGAAACCTTTCACATT 1218
 Qy ProAsnLeuThrValIleThrLeuCysMetPheThrCysPhePheHisLeuThrIleuAsnIle 380
 Db 1219 CCAATATTTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1278
 Qy LeuAlaIleuLeuLeuCysPheGlyAspArgGluPheThrIysAspTrpTrpAsnAlaIys 400
 Db 1279 TTGGCAGAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1338
 Qy SerValGlyAspTrpTrpArgMetTrpAsnMetProValHisIleTyrMetValaArgHis 420

Db 1339 AGTGGGAGATTACTGGAGAAATGTGAATATGCCCTGTTTCATTAATGAGTGTTCGACAT 1398
 Qy IleThrPheProCysLeuArgSerIysIleProIysThrIleAlaIleIleIleAlaPhe 440
 Db 1399 ATATATCTCCCGTGTGTTGCGAGCAAGATACCAGAAACATCTGCCATTTATCTTCTTTC 1458
 Qy LeuValSerAlaValaPheHisGluLeuCysIleAlaValProCysArgLeuPheIysLeu 460
 Db 1459 CTAGCTCTGCAAGCTTTCATGAGCATGATGATGATGATGATGATGATGATGATGATGATGAT 1518
 Qy TrpAlaPheLeuGlyIleMetPheGlnValProleuValIlePheIleThrAsnTyrIleGln 480
 Db 1519 TGGCTTTTCTTGGGATTAATGTTTCAGGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 1578
 Qy GluArgPheGlySerThrValGlyAsnMetIlePheThrPheIlePheCysIlePheGly 500
 Db 1579 GAAAGTTTGGCTCAACGCTGGGGAACATGATCTTCTGTTCACTTCTGATTTTCGGA 1638
 Qy GlnProMetCysValLeuLeuTyrThrHisAspLeuMetAsnArgIysGlySerMetSer 520
 Db 1639 CAACGATGTGTGCTCTTATATACACAGCCTGATGAAACGGAAGATGATGATGATGATGATGAT 1698
 RESULT 4
 AR227739 LOCUS AR227739 1942 bp DNA linear PAT 20-DEC-2002
 DEFINITION Sequence 1 from patent US 6444876.
 ACCESSION AR227739
 VERSION AR227739.1 GI:27266330
 KEYWORDS
 SOURCE
 ORGANISM
 Unknown.
 Unclassified.
 REFERENCE 1 (bases 1 to 1942)
 AUTHORS Lassner, M. W. and Ruzelinsky, D. M.
 TITLE Acyl CoA: cholesterol acyltransferase related nucleic acid
 JOURNAL Sequences
 PATENT: US 6444876-A 1 03-SEP-2002;
 FEATURES location/Qualifiers
 source 1..1942
 BASE COUNT 461 a 421 c 425 g 635 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 5.06e-222 Length: 1942
 Score: 2771.00 Matches: 520
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 6
 US-09-623-514A-2 (1-520) x AR227739 (1-1942)
 Qy 1 MetalAlaIleuAspSerAlaGlyValThrThrValThrGluAsnGlyGlyGluPhe 20
 Db 237 ATGGCGATTTTGGATTTCTGCTGGCTTACTACGGTACGAGGAGAACGGTGGCGAGAGTTTC 296
 Qy ValAspLeuAspArgLeuArgArgIysSerArgSerAspSerAsnGlyLeuLeu 40
 Db 297 GTGATCTTGATPAGGCTTGTGAGAGGAATCGAGATCGATTCCTTCAACGGACTTCCTT 356
 Qy LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60
 Db 357 CTCTCTGTTCCGATTAATAATTCCTTCCGATATATGTGAGAGCTCCCGCGAGACTTGG 416
 Qy AsparglyleaspservalaValasnaspaspaalaglnclyThrAlaasnleuAlaIleasp 80
 Db 417 GATCGGATGTAATCCGCTTGTAACGATGACCTCAGGGAACAGCCAAATTTGGCCGAGAT 476
 Qy AsnansnglyglyAspAsnansnglyglyArgglyglyIleuIleuValasn 100
 Db 477 AATAACGGTGGTGGCGATTAATACGGTGGGAAGAGCGCGGAGGAAGGAAGGAAGAAC 536


```

Db      841 ACCCTGAGAGCTTGCGATATTTCAGTGGCTCCACATTTGTTATAGCCAACTAT
Qy      301 ProArgerSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIle
Db      901 CCACGTTGCGATGATATAGGAAGGTTGGCGCTCGCATATTCGCAAACTGGGCATTA
Qy      321 PheHrgLysPheMetGlyPheIleIleGlnTrpIleAsnProIleValArgAsnSer
Db      961 TTCACCGGATTCATGGATTTATTAATAGAACATATATATATATATATATATATATAT
Qy      341 LysHisProLeuLysGlyAspLeuLeuTrpAlaIleGlnArgValLeuLysLeuSerVal
Db      1021 AAGCATCTCTTAAAGGCGATCTCTATATGCTATTGAAAGAGTGTGAAGCTTTCAGATT
Qy      361 ProAsnLeuTrpValTrpLeuCysMetPheTrpCysPhePheHisLeuTrpLeuAsnIle
Db      1081 CCAATTTTATATGTTGGCTGCGCATGTTCTACTGCTTCTCCACCTTTGGTTAAACATTA
Qy      381 LeuAlaGlnLeuLeuCysPheGlyAspArgGlnPheTrpLysAspTrpTrpAsnAlaLys
Db      1141 TTGGAGAGACTTCTCTGCTTCCGGGATCGTGAATTCACAAAGATTGGATGCAATA
Qy      401 SerValGlyAspTrpTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis
Db      1201 AGTGTGGAGATTACTGGAAGATGGGAATATGCCGTTCATTAATAGATGGTTCGACAT
Qy      421 IleTrpPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIleAlaPhe
Db      1261 AATATCTTCCGCTGCTGCGGCGACGAAGATACCAAGACACCTGCCATTTATCTGCTTTC
Qy      441 LeuValSerAlaValAlaPheHisGlnLeuCysIleAlaValProCysArgLeuPheLysLeu
Db      1321 CTAGTCTCTGCGAGTCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT
Qy      461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIlePheAsnTrpLeuGln
Db      1381 TGGGCTTTCTTGGGATTAATGTTTCAAGGCTTGGCTTTCATCAACAACATCTACAG
Qy      481 GlnArgPheGlySerThrValAlaGlnAsnMetIlePheTrpPheIlePheCysIlePheGly
Db      1441 GAAAGTTTGGCTCAACGCTGGGGAACATGATCTTCTGTTATCTCTGCGATTTTTCGA
Qy      501 GlnProMetCysValLeuLeuTrpTrpHisAspLeuMetAsnArgLysGlySerMetSer
Db      1501 CAACCGATGTGTGCTCTTCTTATATACGACGATGATGAACGAAAGATCGATGTCA
RESULT 2
AF051849      1845 bp      mRNA      linear      PLN 24-JAN-2000
LOCUS      Arabidopsis thaliana diacylglycerol acyltransferase (DAGAT) mRNA,
DEFINITION      complete cds.
ACCESSION      AF051849 AF061563
VERSION      AF051849.1 GI:6625552
KEYWORDS
SOURCE
ORGANISM      Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
1 (bases 1 to 1845)
Boulier-Nave, P., Benveniste, P., Oelkers, P., Sturley, S.L. and
Schaller, H.
Expression in yeast and tobacco of plant cDNAs encoding acyl
CoA:diacylglycerol acyltransferase
Eur. J. Biochem. 267. (1), 85-96 (2000)
JOURNAL
MEDLINE
PUBMED      10601854
2 (bases 1 to 1845)
Benveniste, P.
REFERENCE
AUTHORS      Direct Submission
TITLE      Submitted (28-FEB-1998) Institut de Biologie Moléculaire des
JOURNAL      Plantes, Centre National de la Recherche Scientifique, 28 rue

```

```

COMMENT      Goethe, Strasbourg 67083, France
FEATURES
Location/Qualifiers
1..1845
source
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/cultivar="Columbia"
/db_xref="taxon:3702"
/tissue_type="hypocotyl"
/clone_id="ABRC library, catalog number CD4-15"
/dev_stage="3 day-old seedling"
/notes="screened with EST, GenBank Accession Number
AA042298"
1..1845
/gene="DAGAT"
88..1650
/gene="DAGAT"
/function="catalyzes the acylation of diacylglycerol by
oleoyl-CoA to give triacylglycerol"
/notes="acyl-CoA cholesterol acyltransferase homolog"
/codon_start=1
/product="diacylglycerol acyltransferase"
/protein_id="AAFI9262.1"
/db_xref="GI:6625553"
/translation="MALIDSACVTVTYENGGEFEVDLRLRRKRSRSDSNGLISGS
DNNSPSDDVAPADVRIDSVNDADGTALADNNGCGNNGCGGCGGCGGNAD
ATFTYRSPVAPARRARRESIDSLDFKSHGLFYLQVYLVAVSRILEMLKYGW
LIRDFWFSRSLRDMPLFMCCISLSTPLAFTYKLYVYKHTSIDISLNAADKANPEV
TEVLYPVVTLRCDSDAPLSGVTMLTCTICWAKLYSVAHTSIDISLNAADKANPEV
STYLSKSLATFWAPLTCTOPSTYRACITRKGRWAPKLVITFGMGLTEYIN
PLVNSKHPKGLDLYALIERVAKLSVPLNYWLMCFEFLMLMLIAELLCFQREF
YKDWNNASVGDVPMNMNMPVHKMWVRHYFPCLSKIPKTLAIIIAELVSAVFHELC
IAVPCREKLMALFIMFOVPLVFTNYLQERFGSTGVNMIFMFIQFGPMCVILY
YHDLNRRGSMNS"
BASE COUNT      453 a      378 c      421 g      593 t
ORIGIN
Alignment Scores:
Pred. No.:      4.72e-222      Length:      1845
Score:      2771.00      Matches:      520
Percent Similarity:      100.00%      Conservative:      0
Best Local Similarity:      100.00%      Mismatches:      0
Query Match:      100.00%      Indels:      0
DB:      Gaps:      0
US-09-623-514a-2 (1-520) x AF051849 (1-1845)
Qy      1 MetAlaIleLeuAspSerAlaGlyValThrValThrGlnAsnGlyGlyGlyLeuPhe
Db      88 ATGGCGATTTTGGATTCTGCTGCGCTTACTACGGACGACGAGTGGCGAGAGTTTC
Qy      21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu
Db      148 GTCGATCTTGATAGGCTTCTGCGGAATCGAGATCGGATCTTCTTACCGACTTCTT
Qy      41 LeuSerGlySerAspAsnSerProSerAspAspValGlyAlaProAlaAspValArg
Db      208 CTCTCTGGTTCGGAATATATCTCTCTGATGATGTTGAGCGTCCGCGCAGCTTAGG
Qy      61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp
Db      268 GATCGGATGATTCGCTGTTACGATGACGCTCAAGGACCAATTTGGCCGAGAT
Qy      81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyArgGlyAsn
Db      328 AATAACGGTGGCGGATATATACCGTGTGGAAGAGGCGGAGAGAGAGAGAAC
Qy      101 AlaAspAlaThrPheThrTyraArgProSerValProAlaHisArgArgAlaArgLysLeu
Db      388 GCCGATGCTAGCTTACGATGACGCGTCCAGCTGCATCGGAGGCGGAGAGAGAGT
Qy      121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal

```


BT008883
LOCUS 1563 bp mRNA linear PLN 13-JUN-2003
DEFINITION Arabidopsis thaliana At2g19450 mRNA, complete cds.
ACCESSION BT008883
VERSION BT008883.1 GI:31711931
KEYWORDS FLI CDNA
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 1563)
AUTHORS Kim, C.J., Chen, H., Cheuk, R., Shinn, P., Bowser, L., Carninci, P., Dale, J.M., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Toriumi, M., Wong, C., Wu, H.C., Yamada, K., Yu, G., Shinozaki, K., Davis, R.W., Theologis, A., and Ecker, J.R.
TITLE Arabidopsis ORF clones
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1563)
AUTHORS Kim, C.J., Chen, H., Cheuk, R., Shinn, P., Bowser, L., Carninci, P., Dale, J.M., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Karlin-Neumann, G., Kawai, J., Lam, B., Lin, J., Miranda, M., Narusaka, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Sakurai, T., Satou, M., Seki, M., Southwick, A., Toriumi, M., Wong, C., Wu, H.C., Yamada, K., Yu, G., Shinozaki, K., Davis, R.W., Theologis, A., and Ecker, J.R.
TITLE Direct Submission
JOURNAL Submitted (13-JUN-2003) Salk Institute Genomic Analysis Laboratory (SIGAL), Plant Biology Laboratory, The Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
COMMENT
RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAPL cDNAs (RAPL cDNA: 'RIKEN Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J., Hayashizaki, Y., and Shinozaki, K.
The Salk, Stanford, PGSC (SSP) Consortium members constructed and sequenced the PUNI (ORF) clones using the RAPL cDNAs: Kim, C.J., Chen, H., Cheuk, R., Shinn, P., Bowser, L., Chan, M.M., Chang, C.M., Dale, J.M., Jones, T., Karlin-Neumann, G., Lam, B., Lee, J.M., Lin, J., Miranda, M., Nguyen, M., Onodera, C.S., Palm, C.J., Quach, H.L., Southwick, A., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yamada, K., Yu, G., Davis, R.W., Theologis, A., and Ecker, J.R.
Kim, C.J. (SSP/Salk) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Ecker, J.R. (SSP/Salk) contributed equally to this work as PIs.
LOCATION/Qualifiers
1. 1563
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="2"
/clone="U16373"
/note="This clone is in PUNI 51
ecotype: Columbia"
1. 1563
/note="dlaacylglycerol O-acyltransferase"
/codon_start=1
/product="At2g19450"
/protein_id="AAP68322.1"
/db_xref="GI:31711932"
/translation="MALIDSAGVTYTTENGGEFVVDLRIRKRSKSDSSNGLLSGS
DNPSDDVGAPADVRDRIDSVNDDQGANLANGGDNGGNGGEGEGEGGNAD
LPTTFPSVPAHRRARSPSSDAIFKQSHAFNLCVYLIVNSRLILEMLKGM
LIRTDWFSRSRRLDMPFCISLISFLPAFTVERKLVOKLISEVVLFIHIIIM
TEVLVYVTLRCDSAFLSGVTLMLTLCIKVIAVSHARSHYIFRSIADANDEYV
SYVSLKSLAYFWAPLTCOPSPRSACIRKGMVAROFKATYIIFGAGKTEIOYIN
PIYRNSKHPLKGLDLAIEVYLVLSVFNLYVLCMFYCFPHMLNLAELICGDRKF

YKDMNNAKSVGYDWRMKNMPEVHKMVRNHIYFPLRSKIPRTALIIAFLVSAVEHLC
IAVCRFLKMAFLGIMFVPLVFTNYLQERGSYVNMIFEFICIFQPMCVLLY
YHDLNMRKGSMS"
BASE COUNT 376 a 322 c 371 g 494 t
ORIGIN
Alignment Scores:
Pred. No.: 3,78e-222 Length: 1563
Score: 2771.00 Matches: 520
Percent Similarity: 100.00% Conservative: 0
Best local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0
US-09-623-514a-2 (1-520) x BT008883 (1-1563)
QY 1 MetAlaIleleuSpSerAlaGlyValThThrValThrcLunsnGlyGlyGluPhe 20
DB 1 ATGGCGATTTTGGATTTCTGCGGCTTACGAGGTGACGAGAACGGTGGCGAGATTTC 60
QY 21 ValAspLeuAspArgLeuArgArgGlySerArgSerAspSerSerAsnGlyLeuLeu 40
DB 61 GTGATCTTATAGCGCTTCGTCACGCAATCCAGATCGGATTTCTTAACGACTTCTT 120
QY 41 LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60
DB 121 CTTCTCTGTTCCATATATATATTTCTCCCTTCGATGATGTGGACTCCCGCGAGTTAG 180
QY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
DB 181 GATCGGATTAATTCGTTGTTACGATGAGCGTCAGGAAACGCAATTTGGCGGAGAT 240
QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyValArgGlyGlyGlyGlyGlyGlyGly 100
DB 241 AATTAACGCTGCTGCGCATATTAACGCTGCTGCGAAGACCGCGGAGAGAACGAAAC 300
QY 101 AlaAspAlaThrPheThrTyraArgProSerValProAlaHisArgAlaArgGluSer 120
DB 301 GCCGATCTCAACGATTTGCTGATTCACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 360
QY 121 ProLeuSerSerAspAlaIlePheGlyGlnSerHisAlaGlyLeuPheAsnLeuGlyVal 140
DB 361 CCAGTTGCTGCGACGCAATCTTCMAACAGACGATCCGATTAATCAACCTCTGTGTA 420
QY 141 ValValLeuLeuAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrglyTrp 160
DB 421 GTAGTTCTTATGCTGTAACAGACGATCATCATCAATAATCTTATGAAGTATGCTGC 480
QY 161 LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180
DB 481 TTGATCAGAACGAGATTTCTGTTAGTTCAGATCCGTCGAGATTGGCCGCTTTTCATG 540
QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200
DB 541 TGTGTTATATCCCTTTCGATCTTCTTCTTGCGGCTTACGGGTGAGAAATGGTACTT 600
QY 201 GlnLysTyrlIleSerGluProValValIlePheLeuHisIleIleIleThrMetTrpGlu 220
DB 601 CAGAAATACATATCAGAACCTGTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 660
QY 221 ValLeuTyProValTyValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
DB 661 GTTTTGTATCCACTTACGTCACCCCTAAGCTGATTCCTTTTATACAGCTGCACAG 720
QY 241 LeuMetLeuLeuThrCysIleValITPLeuLysLeuValSerTyrlAlaHisTrpSerTyrl 260
DB 721 TTATGCTCTCTCACTTGATGTGTGGCTAAATGTGTTCTTATGCTCACTTACTACTAT 780
QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProLysValSerTyrlTyrlVal 280
DB 781 GACATTAAGATCCCTAGCAATGACGTGATGAAGCCCAATCTGAAGTCTCTACTACGTT 840
QY 281 SerLeuLysSerLeuAlaTyrlPheMetValAlaProThrLeuGlyTyrlGlnProSerTyrl 300

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 31, 2003, 05:01:50 ; Search time 4675 Seconds
(without alignments)
4550.376 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771
Sequence: 1 MAILDSAGVTVTENGGEF.....OPMCVLYHYHDMNRKGSMS 520

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+pn.model -DEV=xlp
-Q/cgn2.1/USPTO-spool/US9623514/runat.29082003.152159.54/app.query.fasta.1.711
-DB=GenEmbl -QFMT=fastap -SUPERF=range -MINMATCH=0.1 -DOOPCL=0 -DOOEXT=0
-UNIT=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US9623514.ecgn.1.1.5265.ernat.29082003.152159.54 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -NARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.*
1: gb_da:*
2: gb_hg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	1563	8	BT008883
2	2771	100.0	1845	8	AF051849
3	2771	100.0	1904	8	AT238008
4	2771	100.0	1942	6	AR227739
5	2771	100.0	1942	6	AX090345
6	2771	100.0	1942	6	AX090349
7	2771	100.0	1988	8	AT131831
8	2771	100.0	2005	8	AY054480
9	2351.5	84.9	1537	8	AF251794
10	2309.5	83.3	1512	8	AF164434
11	2225	80.3	92822	8	AC005917
12	2225	80.3	97495	8	AC003058
13	1880	67.8	2090	8	AT084032
14	1854	66.9	1964	8	AF298815
15	1842.5	66.1	2099	8	AF129003
16	1833	66.1	1446	8	AF155224
17	813	29.3	43490	8	AP066408
18	794	28.7	1976	9	AF059202
19	784	28.7	1998	9	BC015762
20	793	28.6	2000	9	BC023565
21	791.5	28.6	1518	9	AB057815
22	789	28.5	1856	3	AK114382
23	773.5	27.9	1661	10	AB062763
24	773.5	27.9	1681	10	AB062759
25	773.5	27.9	1681	10	AB062760
26	773.5	27.9	1681	10	AB062761
27	773.5	27.9	1681	10	AB062762
28	773.5	27.9	1766	6	AR227752
29	773.5	27.9	1766	6	AR227753
30	773.5	27.9	1766	6	AX090340
31	772.5	27.9	1792	9	AF236018
32	771.5	27.8	1547	10	AB057816
33	771.5	27.8	1650	10	AF078752
34	771.5	27.8	1776	10	BC003717
35	762.5	27.5	1751	10	AF296131
36	751.5	27.1	1935	4	AY093657
37	735.5	26.5	1493	9	BC006263
38	725.5	26.2	2074	3	AY051835
39	724.5	26.1	1497	3	AF221132
40	720.5	26.0	2117	3	AF468650
41	720.5	26.0	2342	3	AF468649
42	709.5	25.6	1895	6	AR227751
43	709.5	25.6	1895	6	AX090339
44	696	25.1	185932	2	AP003714
45	622.5	22.5	34948	3	CBRG33E23

RESULT 1

ALIGNMENTS


```
Db      640 CGCCCATGCTGGTTATTGGCCCAATGCTTTGCTGGCTGCAATCCAGGTTGAGAA 699
QY      197 sLeuValLeuGlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleTh 217
Db      700 GCGCTGGCGGTGGGTGGCTGACGCGGAGGCGGAGTGGCTGCTGACGTAGCCAACT 759
QY      217 rMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSe 237
Db      760 GCCACCACTTCGTGTTCCAGCGGCTGTGTCTTACTGTTGAGTCTATCACCCTCCAGT 819
QY      237 rGlyValThrLeuMetLeuThr--CysIleValTrpLeuLysLeuValSerTyrAl 256
Db      820 GGGCTCCCTGCTGGCGCTGATGGCGCACACCATCTCTTCTCAAGCTCTTCTCTACCG 879
QY      256 aHisThrSerTyrAspIleArgSer-----LeuAlaAsnAl 268
Db      880 C-----GACGTAACATCATGTGCGCAGGCGCCAGGCCNAGGCTGCCTCTGC 927
QY      268 aAlaAspLysAlaAsnPro-----GluValSerTyr-----TyrValSe 281
Db      928 AGGGAAGAGGAGCGACGCTGCTGCCCGCACACCGTGAAGTACCCGGACAATCTGAC 987
QY      281 rLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrPr 301
Db      988 CTACCGCGATCTCTACTCTCTTCGCCCCACCTTGCTGTGCTACGAGCTCAACTTTC 1047
QY      301 oArgSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIlePhe 321
Db      1048 CGGCTCTCCCGCATCCGGAAGCGCTTCTGTGCGAGGATCCTTGAGATGCTGTCTT 1107
QY      321 eThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSerLy 341
Db      1108 CACCAGCTCCAGGTGGGCTGATCCAGCAGTGGATGCTCCCGCATCCAGACTCCAT 1167
QY      341 sHisProLeuLysGlyAspLeuTyrAla-----IleGluArgValLeuLysLeuSe 359
Db      1168 GAAGCCCTTCAG---GACATGGACTACTCAGCATCATCGAGCGCTCTCTGAAGCTGGC 1224
QY      359 rValProAsnLeuTyrValTrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAs 379
Db      1225 GGTCCCAATCATCCCTGCTCATCTTCTTCTACTGGCTCTCCACTCTCTGCTGAA 1284
QY      379 nIleLeuAlaGluLeuLysCysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAl 399
Db      1285 TGCCGTGGCTGAGTCATGAGTTTGAGACCGGAGTCTACCGGAGCTGGTGGAACTC 1344
QY      399 aLysSerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValAr 419
Db      1345 CGAGTCTGTACCTACTTCTGGCAGAACTGGAACTCCCTGTGCACAAAGTGGTGCATCAG 1404
QY      419 gHisIleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAl 439
Db      1405 ACACCTTACAGCCCATGCTTCGAGGGGAGCAGCAAGTGGATGGCCAGGACAGGGGT 1464
QY      439 aPheLeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLy 459
Db      1465 GTTCTGGCTGGCTTCTTCCAGAGTACGTGGTGGCTGCTCTGCGAATGTTCCG 1524
QY      459 sLeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuVal---PheIleThrAsnTy 478
Db      1525 CCTTGGGCTTTCAGGGCATGATGGCTCAGATCCCACTGGCTGGTGGGCGCGCTT 1584
QY      478 rLeuGlnGluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIl 498
Db      1585 TTTCCAGGCAACTAT-----GGCAACGACGCTGTGTGG---CTGTGCTCAT 1629
QY      498 ePheGlyGlnProMetCysValLeuLeuTyrTyrHisAsp 511
Db      1630 CATCGGACGAGCAATAGCGCTCTCATGTAGCTCCAGCAG 1669
```

QY 366 TrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnTlleLeuAlaGluLeuLeu 385
 Db 391 TGGCTTGGCAGTGTCTAATCTTTTCATTTTGGCTGAACATTTAGCTTAACCTC 450
 QY 386 CysPheGlyAspArgGluPheTyrLysAspTyrPheAsnAlaLysSerValGlyAspTyr 405
 Db 451 TGTTTCGGTCACCGTGAACTCTAATAGGACTGGTGAATGCAAACTGTTGAAGAGTAC 510
 QY 406 TrpArgMetTrpAsnMetProValHisLysTyrMetValArgHisLleTyrPheProCys 425
 Db 511 TGGAGGATGTGGAAACATGCTGTTCATAAGTGGATCATCAGACACATATATTTCATGT 570
 QY 426 LeuArgSerLysIleProLysThrLeuAlaIleIleAlaPheLeuValSerAlaVal 445
 Db 571 ATTAGGNNAGGCTTCCAGGGGTAGCTATTCTAATCTCGTTTGGTTCAGCTGTA 630
 QY 446 PheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGly 465
 Db 631 TTCCATGAGATATGATTGCGGTGCGCGCCACATTTTCAANTCTCGGCATTTCTCGGG 690
 QY 466 IleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySer 485
 Db 691 ATCATGTTTCAGATACCGTTGGTATTCTTGACACAGATATCTCCATGCTACGTTCAAGCAT 750
 QY 486 Thr---ValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCys 504
 Db 751 GTATGTGGGCAACATGATATTGTTGTTTC---TTCAGTATAGTCCGACAGCCGATGNG 807
 QY 505 ValLeuLeuTyrTyrHisAspLeuMet-AsnArgLysGlySerMetSer 520
 Db 808 TGTCTCTATAACTAACAATGACGTGATGAACAAAGCAAGCCCAAGT 856
 RESULT 15
 ID AAA76169 standard; DNA; 1976 BP.
 AC AAA76169;
 XX
 DT 14-DEC-2000 (first entry)
 DE Human ACAT Related Gene Product 1 ARGP1 coding sequence.
 KW Human; ACAT Related Gene Product 1; ARGP1; gene therapy; enzyme;
 KW acyl Coenzyme A-cholesterol acyltransferase 1; ACAT1;
 KW sterol esterification; lipid homeostasis; diacylglycerol acyltransferase;
 KW DGAT; cholesterol; triglyceride biosynthesis; hypertriglyceridaemia;
 KW hyperlipidaemia; atherosclerosis; heart disease; obesity; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 245..1711
 FT /*tag= a
 FT /product= "Human ARGP1"
 XX
 PN US6100077-A.
 XX
 PD 08-AUG-2000.
 XX
 PF 01-OCT-1998; 98US-0165042.
 XX
 PR 01-OCT-1998; 98US-0165042.
 XX
 PA (UYCO) UNIV COLUMBIA NEW YORK.
 XX
 PI Sturley SL, Oelkers P;
 XX
 DR WPI; 2000-557622/51.
 DR P-PSDB; AAB15200.
 XX
 PT New nucleic acid encoding a human diacylglycerol acyltransferase,
 PT useful for treating hyperlipidaemia, atherosclerosis, heart disease, or
 PT other diseases associated with an imbalance of triglyceride levels -

xx Claim 4; Fig 1B; 32pp; English.

xx The enzyme acyl Coenzyme A-cholesterol acyltransferase 1 (ACAT1) mediates
 xx sterol esterification, an important component of intracellular lipid
 xx homeostasis. The present sequence is the coding sequence of human ACAT
 xx Related gene Product 1 (ARGP1). The enzyme encoded by the present
 xx sequence is a diacylglycerol acyltransferase (DGAT). ARGP1 does not
 xx esterify cholesterol. It is thought therefore that ARGP1 participates in
 xx the Coenzyme A-dependent acylation of substrate(s) other than cholesterol
 xx e.g. diacylglycerol. Also, ARGP1 has a predicted diacylglycerol binding
 xx motif, suggesting that it may perform the last acylation in triglyceride
 xx biosynthesis. ARGP1 gene and protein are useful for treating a subject
 xx who has an imbalance in triglyceride levels due to a defect in
 xx esterification of diacylglycerol, via gene therapy. Particularly, ARGP1 is
 xx useful for treating hypertriglyceridaemia, hyperlipidaemia,
 xx atherosclerosis, heart disease, obesity or other diseases associated with
 xx high or excessive levels of triglyceride.

xx Sequence 1976 BP; 323 A; 539 C; 616 G; 398 T; 0 other;

Alignment Scores:

Pred. No.: 1.53e-65 Length: 1976
 Score: 794.00 Matches: 205
 Percent Similarity: 52.53% Conservative: 86
 Best Local Similarity: 37.00% Mismatches: 190
 Query Match: 28.65% Indels: 73
 DB: 21 Gaps: 20

US-09-623-514A-2 (1-520) x AAA76169 (1-1976)

QY 4 LeuAspSerAlaGly-----ValThrThrValThrGluAsnGlyGlyGluPheVal 21
 Db 91 GTTGAACGCGCTGTGTAGCGGCTCACCGGGCTACGCGCGCGCGGCGGCGAGTGGCGG 150
 QY 22 AspLeuAspArgLeuArgArg-----ArgLysSerArgSerAspSerSer 36
 Db 151 CCGTTGTCTAGGCGCGGAGGTGGGCGCGGCTCGGCGGTACGAACCGCGAGGCC 210
 QY 37 AsnGlyLeuLeuLeuSerGlySer----- 44
 Db 211 CACGCTTGGCTGCGCGCGGCTGAGCCATGCGCGCGCGCGCGCGAGTCCCGCGG 270
 QY 45 -----Asp-AsnAsnSerProSerAsp-----AspValGlyAlaProAlaAspValArgAs 61
 Db 271 CCGGAGGACAGGTGCGCGGCTCGAGCCACGCGCGCGGCGGCTCGCGCGCGGAAGA 330
 QY 61 pArgIleAspSerValValAsnAspAspAlaGlnGly---ThrAlaAsnLeuAlaGlyAs 80
 Db 331 AGAG-----GTGCGGAGCGCGCTGCGGGCGCGCGCGCGCGCGGAGCCCGCGGGA 378
 QY 80 pAsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyArgGlyAs 100
 Db 379 CGGCGCGAGCGG 438
 QY 100 nAlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgAlaArgGluSe 120
 Db 439 GCTGAGGTGC-----CATCGCTGCAGGATTC 465
 QY 120 rProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVa 140
 Db 466 TTTATTTCAGCTCTGACATGCGCGGTATTCTTCTGAGAACCTCATCAAGTATGTCAT 522
 QY 140 lValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTr 160
 Db 523 GGTGATGCTCATCTTGAGCAATGCGCGGTATTCTTCTGAGAACCTCATCAAGTATGTCAT 582
 QY 160 pLeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAsp-----TrpPr 177
 Db 583 CTTGGTG---GACCCCATCCAGGTGGTTCTCTGTTCTTCTGAAGGATCCCATAGCTGCC 639
 QY 177 oLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLy 197

```

Db 361 AACTCTGGAGTCAACATACCTTCAAGATGGCTCCGACACTCTGTTATCAAGCCAAAGT 420
Qy 300 TyrPro-ArgSerAlaCysIleArgLysGlyTyrValAlaArgGlnPheAlaLysLeuVa 319
Db 421 TATCTTCGACACCTTATGTAGAAAGGTGGTGGTCCGTCAGGTATCTATATCT 480
Qy 319 lilePheThrGlyPheMetGlyPheIleleuGlnTyrIleAsnProIleValArgAs 339
Db 481 GATATTTACTGCTCCCAAGGATTCATTATTGAGCAATACATAAATCCTATTGTTGTGA 540
Qy 339 nSerLysHisProLeuLysGlyAspLeuTyrAlaIleGluArgValLeuLysLeuSe 359
Db 541 CTCTCAACATCCATTGATGGGAGGATTACTGAATCTGTAGAGACTGTTTGAAGCTCTC 600
Qy 359 rValProAsnLeuTyrValTyrLeuCysMetPheTyrCysPhePheHisLeuTyrLeuAs 379
Db 601 ATTACCAATGTCACCTGCTGGCTTGGCATGTTTATTGCGCTTTCCATCTGTGTTTAA 660
Qy 379 nileLeuAlaGluLeuLysCysPheGlyAspArgGluPheTyrLysAspTyrTrpAsnAl 399
Db 661 CATACTTCTGAGATCTTCGATTGTTGGTGACGAGAAATCTACAAAGACTGGTGCAATGC 720
Qy 399 alySerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValar 419
Db 721 AAAGCAATATGATGATGACTGAGAAATGGAACATGCTGTGCATAAATGGATGTTGCG 780
Qy 419 gHisIleTyrPheProCysLeuArgSerLysLysIleProLysThrLeuAlaIleIleAl 439
Db 781 TCATATATATTCCTTCGATCGGAAATGGTATATCAAGGAAGTGTGCTGTTTTTATATC 840
Qy 439 aPheLeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLy 459
Db 841 GTTCTTTTCTGCTGCTGACTTCATGAGTATGTTGCTGTGCTGCTGCCACATACACTCA 900
Qy 459 sleuTrpIlePheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrIle 479
Db 901 GTTCTGGGCTTCTTAGGAATCATGCTTCAGATTCCTCCCTCATCATATTGACATCATACCT 960
Qy 479 uGlnGluArgPheGlySerThr---ValGlyAsnMetIlePheTrpPheIlePheCysIle 498
Db 961 CAAATAAATTCAGTGACACAAATGGTGGCAATATGATCTTTGGTGTGTTTTCTGCTGAT 1020
Qy 498 ePheGlyGlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArg 515
Db 1021 ATACGGCAGCAATGTGTGTTCTATTGTTATGTTATGATGATGATGATGATGATGATG 1072

RESULT 14
AAA48936
ID AAA48936 standard; cDNA; 901 BP.
XX
AC AAA48936;
XX
DT 06-DEC-2000 (first entry)
XX
DE Corn diacylglycerol acyltransferase cDNA #4.
XX
KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW triacylglycerol; herbicide; EC2.3.1.20; ss.
XX
OS Zea mays.
XX
FH Key
FT 1..858 Location/Qualifiers
FT CDS
FT /tag= a
FT /partial
FT /transl_except= (pos:442..444,aa:Xaa)
FT /transl_except= (pos:490..492,aa:Xaa)
FT /transl_except= (pos:587..589,aa:Xaa)
FT /transl_except= (pos:805..807,aa:Xaa)
FT /transl_except= (pos:820..822,aa:Xaa)
FT /note= "Xaa= unknown"
FT /product= Partial_diacylglycerol_acyltransferase

```

```

XX WO200032756-A2.
PN 08-JUN-2000.
XX
PF 01-DEC-1999; 99WO-US28354.
XX
PR 02-DEC-1998; 98US-0110602.
PR 31-MAR-1999; 99US-0127111.
XX
PA (DUPO ) DU PONT DE NEMOURS & CO E. I.
XX
PI Cahoon EB, Kinney AJ, Cahoon RE;
XX
DR WPI; 2000-412308/35.
DR P-PSDB; AAY94516.
XX
PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for
PT synthesis of triacylglycerols and increasing the level of oils in plant
PT seeds -
XX
PS Claim 18; Page 46; 62pp; English.
XX
CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,
CC soybean and wheat were screened for sequences with homology to a
CC putative acyl CoA cholesterol acyltransferase related gene from
CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
CC sapiens and Mus musculus. The cDNA clones identified from this process
CC were used to form complete diacylglycerol acyltransferase cDNA
CC sequences. The present sequence is corn diacylglycerol acyltransferase
CC cDNA from a contig of clones p0042.cspaf49r, p0122.ckm57r and
CC p0125.czaa61rb. Diacylglycerol acyltransferases are involved in the
CC synthesis of triacylglycerols. Alteration of the expression of the
CC diacylglycerol acyltransferase DNA can be useful for increasing the
CC level of oils in plant seeds. Inhibitors of diacylglycerol
CC acyltransferase may be useful as herbicides.
XX
SQ Sequence 901 BP; 256 A; 172 C; 193 G; 276 T; 4 other;

```

```

Alignment Scores:
Pred. No.: 2,14e-88 Length: 901
Score: 1028.50 Matches: 188
Percent Similarity: 79.06% Conservative: 31
Best Local Similarity: 67.87% Mismatches: 48
Query Match: 37.12% Indels: 10
DB: 21 Gaps: 3

US-09-623-514A-2 (1-520) x AAA48936 (1-901)
Qy 253 ValSerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspLys--- 271
Db 31 GTCCTTATGACATACATAAAATATATGATAGGGTATGTCCTCAAAAGTACTGAGAGGCT 90
Qy 272 -----AlaAsnProGluValSerTyrTyrValSerLeuLysSerLeu 285
Db 91 GCTGCATATGGAATATTATGTCGATCTGAGAAATATGAAAGATCAACCTTTTAAAGTCFA 150
Qy 286 AlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCys 305
Db 151 GTGTACTTCATGTTGGCCCCAACACTTTGTTACCAGCAACACTTATCTCTCAAACTCATGT 210
Qy 306 IleArgLysGlyTyrPheValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMet 325
Db 211 ATTAGAAAGGTTGGGTGACCCAGCACTATAAAGTCGGTGTGTTTACAGCGCTTGATG 270
Qy 326 GlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLys 345
Db 271 GGCCTTCATAATGAGCAATATATAAACCACAAATGTGAAGAATTCCAAAACATCCACTGAA 330
Qy 346 GlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrVal 365
Db 331 GGGAAATTTTGTGATGCTATAGAAAGAGTCTTTAAACACTCTCAGTGCCCAACATTTATGTA 390

```

Db 601 TCATTACCAAAATGCTACCTGGCTTTCATGTTTATTCCTTTCCATCTGTGGTTA 660
 QY 379 AsnIleuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpAsn 398
 Db 661 AACATACCTGCTGAGATCTTCGATTGGTGGAGAGAAATCTACAAAGACTGGTGAAT 720
 QY 399 AlaLysSerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetVal 418
 Db 721 GCAAGACAATGATGAGTACTGGAGAAATGGACATGCTGTGCATAAATGGATTGT 780
 QY 419 ArgHisIleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIle 438
 Db 781 CGTCATATATATTTCTTGCATGCGAATGGTATATCAAGGAAGTCTGTTTATTA 840
 QY 439 AlaPheLeuValSerAlaVal-Phe----- 446
 Db 841 TCCTTCTTGTCTCTGCTACTTTCATGAGTAACCTAATTTACTTTTTCACCTCTTCATCT 900
 QY 446 ----- 446
 Db 901 GCATATATTAAATATATAGTCTCTATTTTCAAAATGTCTTTCGAGTTTCGACATGCT 960
 QY 446 ----- 446
 Db 961 TTGTTCAAACTTACCAGCTGTAGATTACTTGCATGAAGTGCCTATATAAAATTCAA 1020
 QY 447 -----HisI 448
 Db 1021 TTTCAAAATCCAGTCCCTTTCCAGAAAATATGATACATTTTGTGTCATTTGTACACCA 1080
 QY 448 uLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetPh 468
 Db 1081 GTATCGGTGAGTTCCTCCCTCCACATATCAAGTCTGGCTTCTTAGGAATCATGCT 1140
 QY 468 eGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySerThr---Va 487
 Db 1141 TCAGATTCCTCATCATATATGATCATCATACCTCAAAATATAATTCAGTCACACATGGT 1200
 QY 487 lGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCysValLeuLe 507
 Db 1201 TGCAATATGATCTTTGTTTCTGTCATATACGGCAGCCCAATGTGTCTATT 1260
 QY 507 uTyrTyrHisAspLeuMetAsnArg 515
 Db 1261 GTATTACCATGATGATGAACCGG 1285

RESULT 13

AAA48933
 ID AAA48933 standard; cDNA; 1281 BP.
 XX
 AC AAA48933;
 XX
 DT 06-DEC-2000 (first entry)
 XX
 DE Corn diacylglycerol acyltransferase cDNA #1.
 XX
 KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
 KW triacylglycerol; herbicide; EC2.3.1.20; ss.
 XX
 OS Zea mays.
 XX
 FH Location/Qualifiers
 FX 1..1090
 FT CDS
 FT /tag- a
 FT /partial
 FT /transl_except= (pos:283..286,aa:Ile)
 FT /note="This codon has an apparent 1 nucleotide
 FT insertion which alters the reading frame"
 FT /transl_except= (pos:356..358,aa:xaa)
 FT /note="xaa= unknown"
 FT /transl_except= (pos:869..889,aa:YVLLFL)
 FT /product= Partial_diacylglycerol_acyltransferase
 XX

PN WO200032756-A2.
 XX
 PD 08-JUN-2000.
 XX
 PF 01-DEC-1999; 99WO-US28354.
 XX
 PR 02-DEC-1998; 98US-0110602.
 PR 31-MAR-1999; 99US-0127111.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 PA
 PI Cahoon EB, Kinney AJ, Cahoon RE;
 XX
 DR WPI; 2000-412308/35.
 DR P-PSDB; AAY94513.
 XX
 PT Polynucleotides encoding diacylglycerol acyltransferase, useful for
 PT synthesis of triacylglycerols and increasing the level of oils in plant
 PT seeds
 XX
 Claim 15; Page 40-42; 62pp; English.
 XX
 CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,
 CC soybean and wheat were screened for sequences with homology to a
 CC putative acyl CoA cholesterol acyltransferase related gene from
 CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
 CC sapiens and Mus musculus. The cDNA clones identified from this process
 CC were used to form complete diacylglycerol acyltransferase cDNA
 CC sequences. The present sequence is corn diacylglycerol acyltransferase
 CC cDNA from a contig of clones cpjlc.pk005.h23, cen3n.pk0010.c10,
 CC ccol.pk0029.b6. Diacylglycerol acyltransferases are involved in the
 CC synthesis of triacylglycerols. Alteration of the expression of the
 CC diacylglycerol acyltransferase DNA can be useful for increasing the
 CC level of oils in plant seeds. Inhibitors of diacylglycerol
 CC acyltransferase may be useful as herbicides.
 XX
 SQ Sequence 1281 BP; 351 A; 256 C; 245 G; 417 T; 12 other;

Alignment Scores:
 Pred. No.: 2,98e-110 Length: 1281
 Score: 1260.00 Matches: 229
 Percent Similarity: 76.94% Conservative: 48
 Best Local Similarity: 63.61% Mismatches: 71
 Query Match: 45.47% Indels: 12
 DB: 21 Gaps: 2

US-09-623-514A-2 (1-520) x AAA48933 (1-1281)

QY 168 PheSerSerArgSerLeuArgAspTrpProLeuPheMetCysCysIleSerLeuSerIle 187
 Db 1 TTTAATGCTACATCATTCGAGAGCTGCCACTGCTAATGTGTGCTTAGCTACCCATA 60
 QY 188 PheProLeuAlaAlaPheThrValGluLysLeuValLeuGlnLysTyrIleSerGluPro 207
 Db 61 TTTCCCTTGGTGCATTTGCAGTCGAAAGTGGCATTCACAAATCTCGTTAGTATCCT 120
 QY 208 ValValIlePheLeuHisIleIleIleThrMetThrGluValLeuTyrProValTyrVal 227
 Db 121 GCTACTACCTGTTTTCACATCCTTTTACAAACATTTGAATTTGAAATTTATATTCAGTCTGTG 180
 QY 228 ThrLeuArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeuThrCysIle 247
 Db 181 ATTTAAGTGTGATTTCTGCAGTTTATCAGGCTTTGTGTGTGTATTTATTTGCTGCATT 240
 QY 248 ValTrpLeuLysLeuValSerTyrAlaHisThrSerTyrAspIleArg-SerLeuAlaAs 267
 Db 241 GTTTGGCTGAAGCTTGTATCTTTTGCATACAAACCATGATATACGAAATGATC 300
 QY 267 nAlaAlaAspLysAlaAsnProGluValSer-----TyrTyrVa 280
 Db 301 AAGCGGCAAGAGGTGTGATAAATGAAGTACCGCGCTGGCATAGATAATTTACAANCTCC 360
 QY 280 lSerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGln-Pro-Ser 299

Db	1020	ATCGGAATGGTATATCAAGAAGTAAGTCTGTCTTGATATCATCTCTGTCTGCGGTA	1079
Qy	446	PheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGly	465
Db	1080	CTCATGAGATATGTCTGCTGTTCCTCGCGCATCTCAAGTTCTGGGCATTTAGGA	1139
Qy	466	IleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySer	485
Db	1140	ATAATGCTACAGATCCCTTATTCGATTATTCAGCAGCATACCTCAAAAGTAATAATTCAGAT	1199
Qy	486	Thr---ValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCys	504
Db	1200	AAATGGTTGGCAACATGATATTTGGTCTTTTCTGCATCTATGGCAGCCAATGTGC	1259
Qy	505	ValLeuLeuTyrTyrHisAspLeuMetAsnArg	515
Db	1260	CTTCTCCTGTACTATCATGTGATGAACAGG	1292
RESULT 11			
AAA48942			
ID	AAA48942 standard; cDNA; 1975 BP.		
XX	AAA48942;		
XX			
DT	06-DEC-2000 (first entry)		
XX			
DE	Wheat diacylglycerol acyltransferase cDNA #2.		
XX			
KW	Diacylglycerol acyltransferase; corn; rice; soybean; wheat;		
KW	triacylglycerol; herbicide; EC2.3.1.20; ss.		
OS	Triticum aestivum.		
XX			
FH	Key Location/Qualifiers		
FT	107..1633		
FT	CDS		
FT	/tag= a		
FT	/product= Diacylglycerol_transferase		
XX			
PN	WO200032756-A2.		
XX			
PD	08-JUN-2000.		
XX			
PF	01-DEC-1999; 99WO-US28354.		
XX			
PR	02-DEC-1998; 98US-0110602.		
PR	31-MAR-1999; 99US-0127111.		
XX			
PA	(DUPO) DU PONT DE NEMOURS & CO E I.		
XX			
PI	Cahoon EB, Kinney AJ, Cahoon RE;		
XX			
PI	WPI; 2000-412308/35.		
DR	P-PSDB; AAY94522.		
XX			
PT	Polynucleotides encoding diacylglycerol acetyltransferase, useful for		
PT	synthesis of triacylglycerols and increasing the level of oils in plant		
PT	seeds		
XX			
PS	Claim 4; Page 56; 62pp; English.		
XX			
CC	In the present invention, cDNA libraries from Arabidopsis, corn, rice,		
CC	soybean and wheat were screened for sequences with homology to a		
CC	putative acyl CoA cholesterol acyltransferase related gene from		
CC	Arabidopsis thaliana and diacylglycerol acyltransferases from Homo		
CC	sapiens and Mus musculus. The cDNA clones identified from this process		
CC	were used to form complete diacylglycerol acyltransferase cDNA		
CC	sequences. The present sequence is wheat diacylglycerol		
CC	acyltransferase cDNA from clone wri.pK0119.b6.fis. Diacylglycerol		
CC	acyltransferases are involved in the synthesis of triacylglycerols.		
CC	Alteration of the expression of the diacylglycerol acyltransferase		
CC	DNA can be useful for increasing the level of oils in plant seeds.		
CC	Inhibitors of diacylglycerol acyltransferase may be useful as		
CC	herbicides.		

|||||
 1445 AATAAGATTTTGGTTCATATTCAGTATCCCTGGTCAACCTATGTGTACTGTCTATAC 1504
 QY 509 ThrHisAspLeuMetAsnArgLysGlySerMet 519
 |||||
 Db 1505 TACCATGACTTGATCAATAGGAAGGCAACTT 1537

RESULT 10

AAA48938
 ID AAA48938 standard; cDNA: 1587 BP.

XX
 AC AAA48938;

XX 06-DEC-2000 (first entry)

XX Rice diacylglycerol acyltransferase cDNA #2.

XX Diacylglycerol acyltransferase; corn; rice; soybean; wheat;

XX triacylglycerol; herbicide; EC2.3.1.20; ss.

XX Oryza sativa.

XX Key Location/Qualifiers

XX CDS 15..1310

XX /tag= a

XX /partial

XX /note= "The CDS of this sequence only encodes

XX amino acids 70 to 500 of the protein in AAY94518"

XX WO200032756-A2.

XX 08-JUN-2000.

XX 01-DEC-1999; 99WO-US28354.

XX 02-DEC-1998; 98US-0110602.

XX 31-MAR-1999; 99US-0127111.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Cahoon EB, Kinney AJ, Cahoon RE;

XX WPI; 2000-412308/35.

XX P-PSDB; AAY94518.

XX Polynucleotides encoding diacylglycerol acyltransferase, useful for
 XX synthesis of triacylglycerols and increasing the level of oils in plant
 XX seeds

XX Claim 4; Page 48-49; 62pp; English.

XX In the present invention, cDNA libraries from Arabidopsis, corn, rice,
 XX soybean and wheat were screened for sequences with homology to a
 XX putative acyl CoA cholesterol acyltransferase related gene from
 XX Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
 XX sapiens and Mus musculus. The cDNA clones identified from this process
 XX were used to form complete diacylglycerol acyltransferase cDNA
 XX sequences. The present sequence is rice diacylglycerol acyltransferase
 XX cDNA from clone rls24.pk0034.d8:fls. Diacylglycerol acyltransferases are
 XX involved in the synthesis of triacylglycerols. Alteration of the
 XX expression of the diacylglycerol acyltransferase DNA can be useful for
 XX increasing the level of oils in plant seeds. Inhibitors of
 XX diacylglycerol acyltransferase may be useful as herbicides.

XX Sequence 1587 BP; 411 A; 323 C; 334 G; 519 T; 0 other;

Alignment Scores:

Pred. No.: 2,65e-142 Length: 1587
 Score: 1597.50 Matches: 284
 Percent Similarity: 80.05% Conservative: 61
 Best Local Similarity: 65.89% Mismatches: 73
 Query Match: 57.65% Indels: 13
 DB: 21 Gaps: 3

US-09-623-514a-2 (1-520) x AAA48938 (1-1587)

QY 93 GlyGlyGlyGluGlyArgGlyAsnAlaAspAlaThrPheThrTyrArgProSerValPro 112
 |||||
 Db 15 GCGGTGGGAC-----TTCTCCGGGTTACGTTCCCGCGCGCGCGG 59
 QY 113 AlaHisArgAlaArgGluSerProLeuSerSerAspAlaIlePheLysGlnSerHis 132
 |||||
 Db 60 GTGCACCGCAAGGCCAAGGAGAGCCCTCAGCTCCGAGCCATCTTCAAGCAGAGTCAT 119
 QY 133 AlaGlyLeuPheAsnLeuCysValValValLeuLeuAlaValAsnSerArgLeuIle 152
 |||||
 Db 120 GCAGGCCTTTCAACCTATGCATTGTGTCTAGTTGCAGTGAACGAGGAGGCTTATTATC 179
 QY 153 GluAsnLeuMetLysTyrGlyTrpLeuIleArgThrAspPheTrpPheSerSerArgSer 172
 |||||
 Db 180 GAGAACTTAAGAACTATGGCTTATTAATAAGAGCTGGGTTTGGTTTATGATTAATCA 239
 QY 173 LeuArgAspTrpProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaAla 192
 |||||
 Db 240 TTGCGGAGCTGGCCACTTCTAATGTGTCTTAGTCTGCCGTCTTCCCTCGGTGCA 299
 QY 193 PheThrValGluLysLeuValLeuGlnLysTyrIleSerGluProValValIlePheLeu 212
 |||||
 Db 300 TTTGCAGTTGAAAGTTGGCATTTAACAAATGTTATTACTGATGCTGTTGCTACCTGCCCTC 359
 QY 213 HisIleIleThrMetThrGluValLeuTyrProValTyrValThrLeuArgCysAsp 232
 |||||
 Db 360 CATATCTTCCTTTCAACCAACCGAAATGTATATCCAGTGTGCTGATCTTAAAGTGTGAT 419
 QY 233 SerAlaPheLeuSerGlyValThrLeuMetLeuLeuThrCysIleValTrpLeuLysLeu 252
 |||||
 Db 420 TCTGCAGTTTGTCTGGCTTTTGTGATATTATTTGCTGTATTGTTGGTGAAGCTT 479
 QY 253 ValSerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspLysAla 272
 |||||
 Db 480 GTATCTTTTGCACATACAAACCATGATATAGGCAACTGACCATGGCGGCAAGAAGGTT 539
 QY 273 AsnProGluValSerTyrTyr-----ValSerLeuLysSerLeu 285
 |||||
 Db 540 GATAATGAACCTAAGCACAGTTGACATGGATATTTACAACTCCCACTTATGGGATCTTA 599
 QY 286 AlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCys 305
 |||||
 Db 600 ATATATCTTCATGATGGCTCCCTACACTCTGTATACAGCAAGCATATCCCGAAGTTCATGT 659
 QY 306 IleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMet 325
 |||||
 Db 660 GTTAGAAAAGGTGGCTGATTCGTCAAAATTATCTGTACTTGTGATCTTACTGCTCTTCAA 719
 QY 326 GlyPheIleLeuGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLys 345
 |||||
 Db 720 GCCTTCATATTGACCAATACATAATCAATGTTGTGAATCTCAGATCCATCCATGAAA 779
 QY 346 GlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrVal 365
 |||||
 Db 780 GGAGGACTCCCTAAATGCTGTAGAGACTGTTTGAACCTCTCATTTACCAATGTTTACCTG 839
 QY 366 TrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLeu 385
 |||||
 Db 840 TGGCTTTCATGCTTCTATGCTTTTCCATCTCTGTTTAAAGTATATCTTCTGCTGAGATCTT 899
 QY 386 CysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLysSerValGlyAspTyr 405
 |||||
 Db 900 CGATTTGGTGACCGTGAATCTTACAAAGATTGGTGAATGCAAAACCAATTCATGATGAT 959
 QY 406 TrpArgMetTrpAsnMetProValHisLysTrpMetValArgHisIleTyrPheProCys 425
 |||||
 Db 960 TGGAGAAATGGAATATGCCGTGACATAAATGGGTGTGTCGCATATTTACTTTCTTCTTGC 1019
 QY 426 LeuArgSerLysIleProLysThrLeuAlaIleIleLeuAlaPheLeuValSerAlaVal 445
 |||||

XX	01-DEC-1999;	99WO-US28354.
PF		
XX	02-DEC-1998;	98US-0110602.
XX		
PR	98US-0110602.	
PR	31-MAR-1999;	99US-0127111.
XX		
XX	(DUPO) DU PONT DE NEMOURS & CO E. I.	
PA		
XX		
XX	Cahoon EB, Kinney AJ, Cahoon RE;	
PI		
XX	WPI; 2000-412308/35.	
DR	P-PSDB; ANY94519.	
DR		
XX		
PT	Polynucleotides encoding diacylglycerol acetyltransferase, useful for	
PT	synthesis of triacylglycerols and increasing the level of oils in plant	
PT	seeds -	
XX		
XX	Claim 4; Page 51; 62pp; English.	
XX		
CC	In the present invention, cDNA libraries from Arabidopsis, corn, rice,	
CC	soybean and wheat were screened for sequences with homology to a	
CC	putative acyl CoA cholesterol acyltransferase related gene from	
CC	Arabidopsis thaliana and diacylglycerol acyltransferases from Homo	
CC	sapiens and Mus musculus. The cDNA clones identified from this process	
CC	were used to form complete diacylglycerol acyltransferase cDNA	
CC	sequences. The present sequence is soybean diacylglycerol	
CC	acyltransferase cDNA from clone srl.pk0098.a8. Diacylglycerol	
CC	acyltransferases are involved in the synthesis of triacylglycerols.	
CC	Alteration of the expression of the diacylglycerol acyltransferase	
CC	DNA can be useful for increasing the level of oils in plant seeds.	
CC	Inhibitors of diacylglycerol acyltransferase may be useful as	
CC	herbicides.	
CC		


```
QY 404 ----- 404
Db 3859 TGAATTTAAAGTGTGCGTGTGTTTGGGTCAATCACTAAACAAATTCATGTAATCACT 3918
QY 405 ----- 411
Db 3919 GTCCTCTTATCAGTACTGGAGATGTGGAATATGGTATGGTCTCTCTCCCTAAACATCA 3978
QY 411 ----- 411
Db 3979 CCTTCTTTGTACACAAATAGAGAAGAGAGCTAATTAAGATCTTGTGTTTCTCCTGCACAG 4038
QY 412 ProValHisLysTrpMetValargHisIleTyrPheProCysLeuArgSerLysIlePro 431
Db 4039 CCTGTTCATAAATGGATGTTCCACATATATACCTCCCGTGTGCGCAGAGATACCA 4098
QY 432 Lys----- 432
Db 4099 AA-GGTGAGTGAGATATATACCGATATGCAATGTGCGAGATTTGTTCTGTGATATAAT 4157
QY 433 ----- 443
Db 4158 TTAACCTCCACACACTTGTGTTTCAGACACTCGCCATTATCATTCCTTCTAGTCTCT 4217
QY 444 AlaValPheHisGlu----- 448
Db 4218 GCAGTCTTTCATGAGGTATACATACATTTCTACATTCGCCCTGTCTCTAGACGCATGAACAC 4277
QY 448 ----- 448
Db 4278 ACGCTAGTGAAGAATGCTATATATCAAGCATGTTGTTTACTTAACGATCTGTGTGA 4337
QY 449 ----- 462
Db 4338 CAAATTTCCCTTTGACAGCTATGATCGAGTTCCTTGTCTCTCTCAAGCTATGGCT 4397
QY 463 PheLeuGlyIleMetPheGln----- 469
Db 4398 TTTCTTGGGATTATGTTTCAGGTTAAATAAATTACTAACTGCTGCAGTCGATTTTACTA 4457
QY 470 ----- 475
Db 4458 AACTCTAATCTCATATTCGACCAACCAATTTGTTGAGTAGTGCTTTGGTCTTCATC 4517
QY 476 ThrAsnTyrLeuGlnGluArgPheGlySerThrVal----- 487
Db 4518 ACAAACTATCTACAGGAAGGTTTGGCTCAACGGT-ATGCTCTCAAAACCCGAGAAATA 4576
QY 487 ----- 487
Db 4577 GAACGAATAACTCTTTTCTTTCATAGCTAGCCATTTAAATCGCAATGCTGAAACTTAATA 4636
QY 488 ----- 493
Db 4637 ATAAAGGTGATCTGTTTGGATGGATCATATATATAGTGGGGGAACATGATCTCTCG 4696
QY 494 PheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTyrTyrHisAspLeuMet 513
Db 4697 TTCATCTTCTGCATTTTCGACAAACCGATGTGTGCTCTCTTTATTACCAACCGCTGATG 4756
QY 514 AsnArgLysGlySerMetSer 520
Db 4757 AACCGAAAAGGATCGATGTCA 4777

RESULT 8
AAA51485
ID AAA51485 standard; DNA; 5339 BP.
XX
AC AAA51485;
XX
DT 09-OCT-2000 (first entry)
XX
DE A. thaliana AS11 diacylglycerol acyltransferase genomic DNA.
```

```
XX
KW DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
OS size; weight; carbon flux; TAG1; insertion mutant; ds.
XX Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 1761..4926
FT exon /tag= a
FT exon /tag= b
FT intron /partial
FT exon 2151..2234
FT exon /tag= c
FT exon 2235..2315
FT intron /tag= e
FT intron 2316..2380
FT misc_RNA /tag= f
FT 2370..2515
FT /tag= d
FT exon /note= "insertion region"
FT 2381..2461
FT intron /tag= g
FT /note= "duplication of exon 2"
FT 2462..2552
FT exon /tag= h
FT exon 2553..2622
FT intron /tag= i
FT 2623..2728
FT exon /tag= j
FT 2729..2804
FT intron /tag= k
FT 2805..2901
FT exon /tag= l
FT 2902..2970
FT intron /tag= m
FT 2971..3050
FT exon /tag= n
FT 3051..3174
FT intron /tag= o
FT 3175..3234
FT exon /tag= p
FT 3235..3332
FT intron /tag= q
FT 3333..3410
FT exon /tag= r
FT 3411..3512
FT intron /tag= s
FT 3513..3612
FT exon /tag= t
FT 3613..3749
FT intron /tag= u
FT 3750..3893
FT exon /tag= v
FT 3894..3975
FT intron /tag= w
FT 3976..4079
FT exon /tag= x
FT 4080..4100
FT intron /tag= y
FT 4101..4184
FT exon /tag= z
FT 4185..4247
FT intron /tag= aa
FT 4248..4330
FT exon /tag= ab
FT 4331..4378
FT intron /tag= ac
FT 4379..4501
FT exon /tag= ad
FT 4502..4564
FT intron /tag= ae
FT 4565..4645
```

Best Local Similarity:	51.04%	Mismatches:	5
Query Match:	80.30%	Indels:	493
DB:	21	Gaps:	15
US-09-623-514A-2 (1-520) x AAA51483 (1-5193)			
QY	1	MetAlaIleLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGluPhe	20
DB	1761	ATGGCGATTGTTGATTCTGCTGGCTGTTACTACGGTGACGGAGAACGGTGGCGAGAGTTC	1820
QY	21	ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu	40
DB	1821	GTGATCTTGATAGGCTTCGTCACCGAAATCAGATCGATTCTTCAACGACATCTT	1880
QY	41	LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg	60
DB	1881	CTCTCTGGTCCGATATAATCTCCTTCGGATGATGTTGGAGCTCCCGCGACGTTAGG	1940
QY	61	AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp	80
DB	1941	GATCGGATGATTCCTGTTGTTAACGATGACGCTCAGGGAAACCAATTTGGCGGAGAT	2000
QY	81	AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGluGlyArgGlyAsn	100
DB	2001	AATAACGGTGTGGCGATAATAACGCTGGTGGAGAGCGCGGAGAGAGAGAAAC	2060
QY	101	AlaAspAlaThrPheThrArgProSerValProAlaHisArgAlaArgGluSer	120
DB	2061	GCGGATGCTACGTTTACGTATCGACGCTCGGTTCCAGCTCATCGGAGCGGAGAGAGT	2120
QY	121	ProLeuSerSerAspAlaIlePheLys	129
DB	2121	CCACTTAGCTCCGACGCAATCTTCAACACAGGTTTAAATCTCAGAAATCTCGAATTGG	2180
QY	130	-----GlnSerHis	132
DB	2181	TGTTTCTGCTGTTTATATGAAATGAGTTGGTGGTATGTTTTCATGTCAGACCAT	2240
QY	133	AlaGlyLeuPheAsnLeuCysValValValLeuIleAlaValAsnSerArgLeuIle	152
DB	2241	GCGGATATTACACCTCTGTGTAGTAGTTCTTATTGCTGTTAAACAGTAGACTCATCATC	2300
QY	153	GluAsnLeuMetLys	157
DB	2301	GAAATCTTATGAGGTTTGGCTGTTACTGTTCTCCTTTTAGGAATGGAATGCTGTA	2360
QY	158	-----TyrGlyTrpLeuIle	162
DB	2361	AATTTATCAGACGAGAAATTAATTTGTTGTTGCTATCATTCATGATGTTGGTGTAT	2420
QY	162	eArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet	180
DB	2421	CAGAACGGATTTCTGTTTGTAGTTTCAAGATCGCTGCGAGATTGGCGCTTTTCATGTGTTG	2480
QY	180	-----	180
DB	2481	GTAAGAAGAGATGTTTTTATTTCACGAATGTTACATGTTATACGTAATATGATGAGT	2540
QY	181	-----CysCysIleSerLeuSerIlePhe	188
DB	2541	TTAGTGATCAAGTTCCTTTTGATTCTTCTTCTTCTGCTGCTATATCCCTTTCGATCTT	2600
QY	188	eProLeuAlaAlaPheThrValGluLysLeuValLeuGlnLysTrpIleSerGluPro	207
DB	2601	TCCTTTGGCTGCTTTACGGTTGAGAAATTTGTTACTTTCAGAAATACATATCAGAACCTGT	2660
QY	207	-----	207
DB	2661	GAGTAATTAATTCACGACCATTAATCTGTAATTTTATGAGACAAAGTTTGTATCATG	2720
QY	208	-----ValValIlePheLeuHisIleIleIle	216
DB	2721	AAGAACTTACAGTTCTGTTTGAATAATGCTCAAGGTTGTCATCTTCTCATATATTTA	2780
QY	216	leThrMetThrGluValLeuTyrProValTyrValThrLeu	229
DB	2781	TCACCATGACAGAGGTTTGTATCCAGTTTACGTACACCTTAAGGTGATGACTGTTTCTG	2840
QY	229	-----	229
DB	2841	GTCTCAGTTTGTGATGACTGTTTTTAAAGTTTGTGTGACCGGGTATCTTGAAATAGG	2900
QY	230	--ArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeuThrCysIleVal	249
DB	2901	ACAGGTGTGATCTGCTTTTATFACAGGTGTCTACTTTGATGCTCCCTACCTTGCATGTGT	2960
QY	249	rpLeuLysLeuValSerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAla	269
DB	2961	GGCTAAAGTTGGTTTCTTATGCTCATACTAGCTATGACATAAGATCCCTAGCCAATGAC	3020
QY	269	laAspLysAla	272
DB	3021	CTGATAAGGT-AAAATACGAAAAAGAGCGTATGTATTAGTCACTTGCACTGTGTACTG	3079
QY	273	-----AsnProGluValSerTyrTyrValSerL	282
DB	3080	TTTTAACCAACACTGTTATGAATTTAGGCCAATCCTGAAGTCTCTACTACTAGTACT	3139
QY	282	eulysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyr	296
DB	3140	TGAAGAGCTGGCATATTTTCATGCTGCCACATTTGTTATCAGTAACTGCAAGT	3199
QY	296	-----	296
DB	3200	GCATCAACCACTTATATCTTGTCAAGAGTTTCTGTCTAAACCTCGATCTTTCCTTTTC	3259
QY	297	--GlnProSerTyrProArgSerAlaCysIleArgLysGlyTrpValAlaArgGluPheA	316
DB	3260	CCAGCCAGTATFATCCAGCTTCTGCATGTATACGGAAGGTTGGTGGTCTGCTCAATTTG	3319
QY	316	laLysLeuValIlePheThrGlyPheMetGlyPheIleLeuGlu	330
DB	3320	CMAAAGTGTGATATTCACCGGATTCATGGGATTTATATGAAACAGTACGTTTTCACA	3379
QY	330	-----	330
DB	3380	TCCTGCTTTATATGATTTCTTCTGTTGAAATCATCATCTGCTGCTGCTGCTGCTGCTG	3439
QY	331	-----GlnTyrIleAsnProIleValArgAsnSerLysHis	342
DB	3440	TCATGTTCTTTTGTACATTTTGGCAGTATATAAATCTTATGTCAGGAACTCAAGCAT	3499
QY	343	ProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsn	362
DB	3500	CCTTTGAAAGGGCATCTTCTATATGCTATTGAAAGAGTGTGAAGCTTTCAGTTCCAAAT	3559
QY	363	LeuTyrValTrpLeuCysMetPheTyrCysPhePheHisLeuTrp	377
DB	3560	TTATATGTGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	3619
QY	377	-----	377
DB	3620	TCCTTTTCAAAATAATTTGCAAAATTCGAAAAACCGAAAAAGGCTAAATCTCATACGAAT	3679
QY	377	-----	377
DB	3680	TGATATTTTGTAGTTTCTTAGAGTGGGTGATGTAATTTTACTGTAAGCAAAATCTCTT	3739
QY	378	-----LeuAsnIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLys	394
DB	3740	GTCCAAAGGTAAACATATTTGGCAGAGCTTCTCTGCTTCGGGATCGTGAATTTCTACAA	3799
QY	395	AspTrpTrpAsnAlaLysSerValGlyAsp	404
DB	3800	GATTGTGGAATGCAAAAAGTGTGGGAGA-TGTGAGCTATTTTACTCAAAAGAAAACTTA	3858


```

Db 1459 CATAAATGGATGGTTCGACATATATATCTCCCGTGGCTTGGCAGCAAGATACCAAGACA 1518
Qy 434 LeuAlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLeuCysIleAlaVal 453
Db 1519 CTCGCCATTATCATGCTTCTCTAGTCTCTCAGTCTTTCATGAGCTATGATCGCAGTT 1578
Qy 454 ProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuVal 473
Db 1579 CCTTGTGCTCTCTTCAAGCTATGGCTTTCTTGGGATTATGTTTCAGGTGCCCTTGGTC 1638
Qy 474 PheIleThrAnrTyLeuGlnGluArgPheGlySerThrValGlyAsnMetIlePheTrp 493
Db 1639 TTCATCAAAACTATACAGGAAGGTTTGGCTCAACGGTGGGGAACATGATCTTCFG 1698
Qy 494 PheIlePheCysIlePheGlyGlnPromMetCysValLeuLeuTyTrHisAspLeuMet 513
Db 1699 TTCATCTCTGCATTTTCGGACACCGATGTGTGCTTCTTTATACCGACGCTGATG 1758
Qy 514 AsnArgLysGlySerMetSer 520
Db 1759 AACCGAAAGGATCGATGCA 1779

RESULT 7
AAAS1483
ID AAA51483 standard; DNA; 5193 BP.
AC
XX
XX
XX 09-OCT-2000 (first entry)
XX
XX
XX A. thaliana diacylglycerol acyltransferase genomic DNA.
KW DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW size; weight; carbon flux; ds.
XX
XX Arabidopsis thaliana.
XX
FH Key Location/Qualifiers
FT CDS 1761..4780
FT FT /*tag= a
FT exon 1761..2150
FT FT /*tag= b
FT FT /partial
FT intron 2151..2234
FT FT /*tag= c
FT FT 2223..2369
FT FT /*tag= d
FT FT /note= "region duplicated in insertion mutant AS11
FT FT (see AAA51485)"
FT FT 2235..2315
FT FT /*tag= e
FT intron 2316..2406
FT FT /*tag= f
FT exon 2407..2480
FT FT /*tag= g
FT intron 2481..2582
FT FT /*tag= h
FT exon 2583..2661
FT FT /*tag= i
FT intron 2662..2758
FT FT /*tag= j
FT exon 2759..2823
FT FT /*tag= k
FT intron 2824..2927
FT FT /*tag= l
FT exon 2928..3028
FT FT /*tag= m
FT intron 3029..3108
FT FT /*tag= n
FT exon 3109..3186
FT FT /*tag= o
FT intron 3187..3264
FT FT /*tag= p

```

```

FT exon 3265..3366
FT FT /*tag= q
FT intron 3367..3466
FT FT /*tag= r
FT exon 3467..3603
FT FT /*tag= s
FT intron 3604..3747
FT FT /*tag= t
FT exon 3748..3829
FT FT /*tag= u
FT intron 3830..3933
FT FT /*tag= v
FT exon 3934..3954
FT FT /*tag= w
FT intron 3955..4038
FT FT /*tag= x
FT exon 4039..4101
FT FT /*tag= y
FT intron 4102..4184
FT FT /*tag= z
FT exon 4185..4232
FT FT /*tag= aa
FT intron 4233..4355
FT FT /*tag= ab
FT exon 4356..4418
FT FT /*tag= ac
FT intron 4419..4499
FT FT /*tag= ad
FT exon 4500..4550
FT FT /*tag= ae
FT intron 4551..4675
FT FT /*tag= af
FT exon 4676..4780
FT FT /*tag= ag
FT FT /partial
FT XX
PN WO200036114-A1.
XX
XX 22-JUN-2000.
PD
XX
XX 16-DEC-1999; 99WO-CA01202.
PF
XX
XX 17-DEC-1998; 98US-0112812.
PR
XX
XX (CANA ) NAT RES COUNCIL CANADA.
PA
XX
XX Zou J, Taylor DC, Wei Y, Jako CC;
PI
XX
XX WPI; 2000-431592/37.
DR
XX
XX P-PSDB; AAY96853.
XX
XX New DNA encoding diacylglycerol acyltransferase from Arabidopsis
PT thaliana for transforming plants and regulating seed oil content, fatty
PT acid synthesis and seed oil acyl composition in commercial and crop
PT plants
XX
XX Claim 2; Page 67-68; 91pp; English.
PS
XX
XX This DNA encodes Arabidopsis thaliana diacylglycerol acyltransferase
CC (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating
CC seed oil content, the ratio of diacylglycerol/triacylglycerol proportions
CC in seed oil, fatty acid synthesis, seed oil acyl composition, seed
CC size/weight and carbon flux into other seed components in commercial and
CC crop plants. The natural formation of triacylglycerols can be modified to
CC increase the yield in commercial plant oils or modify their composition
CC to achieve specific commercial improvements of plants and plant products.
XX
XX Sequence 5193 BP; 1582 A; 894 C; 903 G; 1814 T; 0 other;
SQ

```

```

Alignment Scores:
Pred. No.: 2.12e-201 Length: 5193
Score: 2225.00 Matches: 515
Percent Similarity: 51.04% Conservative: 0

```

CDS 139..1782
 /tag- a
 /product- diacylglycerol_acyltransferase
 /note- "Insertion mutant"
 610..690
 /tag- b
 /note- "81 bp insertion, duplication of exon 2"
 WO200036114-A1.
 XX XX
 PD 22-JUN-2000.
 XX XX
 PF 16-DEC-1999; 99WO-CA01202.
 XX XX
 PR 17-DEC-1998; 98US-0112812.
 XX XX
 PA (CANADA) NAT RES COUNCIL CANADA.
 XX XX
 PI Zou J, Taylor DC, Wei Y, Jako CC;
 DR WPI; 2000-431592/37.
 DR P-PSDB; AAY96854.
 XX XX
 PT New DNA encoding diacylglycerol acyltransferase from Arabidopsis
 PT thaliana for transforming plants and regulating seed oil content, fatty
 PT acid synthesis and seed oil acyl composition in commercial and crop
 PT plants
 XX XX
 PS Claim 5; Page 79; 9lpp; English.
 CC CC
 CC This cDNA encodes Arabidopsis thaliana ecotype Columbia mutant AS11
 CC diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a
 CC 147 bp insertion located at the central region of intron 2. The insertion
 CC is a duplication of a segment that is composed of 12 bp from the 3' end
 CC of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the
 CC 5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful
 CC for regulating seed oil content, the ratio of diacylglycerol to
 CC triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil
 CC acyl composition, seed size/weight and carbon flux into other seed
 CC components in commercial and crop plants. The natural formation of
 CC triacylglycerols can be modified to increase the yield in commercial
 CC plant oils or modify their composition to achieve specific commercial
 CC improvements of plants and plant products.
 XX XX
 SQ Sequence 1985 BP; 482 A; 415 C; 446 G; 642 T; 0 other;
 Alignment Scores:
 Pred. No.: 7.89e-252 Length: 1985
 Score: 2747.50 Matches: 520
 Percent Similarity: 95.06% Conservative: 0
 Best Local Similarity: 95.06% Mismatches: 0
 Query Match: 99.15% Indels: 27
 DB: 21 Gaps: 1
 US-09-623-514A-2 (1-520) x AAA51484 (1-1985)
 QY 1 MetAlaIleuAspSerAlaGlyValThrThrValThrGluAsnGlyGlyGluPhe 20
 DB 139 ATGGCGATTGTGATTCTGCTGCGGTTACTACGGTGACGAGAACGGTGGCGGAGAGTTC 198
 QY 21 ValAspLeuAspArgLeuArgArgGlySerArgSerAspSerSerAsnGlyLeuLeu 40
 DB 199 GTCGATCTTGATAGGCTTCGTCACGCGAATCGAGATCGGATCTCTTACAGGACTTCCT 258
 QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60
 DB 259 CTCCTCGGTCCGATAATAATCTCCTCGGATGATGTTGGAGCTCCCGCGGACGTTAGG 318
 QY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
 DB 319 GATCGGATGATTCCTGTTTACGATGACGCTCAGGGAACAGCCCAATTTGGCCGGAGAT 378
 QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyAsn 100

DB 379 AATAACGGTGGTGGCGATAATAACGGTGGTGGAGAGCGCGGAGAGGAGGAAAC 438
 QY 101 AlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgAlaArgGluSer 120
 DB 439 GCCGATGCTACGTTAGGTATCGACCGTGGTTCACGCTCATCGGAGGGCGAGAGAGT 498
 QY 121 ProLeuSerSerAspAlaIlePheLysGln----- 130
 DB 499 CCACCTTAGCTCCGACGCAATCTTCAACAGAGCCATGCCGGATTATTCACACCTCTGTGTA 558
 QY 131 -----SerHisAla 133
 DB 559 GTAGTTCTTATTGCTGTAAACACAGTAGACTCATCATCGAAATCTTATGAAGAGCCATGCC 618
 QY 134 GlyLeuPheAsnLeuCysValValValLeuIleAlaValAsnSerArgLeuIleGlu 153
 DB 619 GGATTATTCAACCTCTGTGTAGTCTTATTGCTGTAAACAGTAGACTCATCATCGAA 678
 QY 154 AsnLeuMetLysTyrGlyTrpLeuIleArgThrAspPheTrpPheSerSerArgSerLeu 173
 DB 679 AATCTTATGAAGTATGGTTGGTTGATCAGAACGATTCTTGGTTTAGTTCAAGATCGCTG 738
 QY 174 ArgAspTrpProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaIlePhe 193
 DB 739 CGAGATTGGCGCTTTTCATGTGTGTATATCCCTTTCGATCTTCTCTTTGGGCTGCTTT 798
 QY 194 ThrValGluLysLeuValLeuGlnLysTyrIleSerGluProValValIlePheLeuHis 213
 DB 799 ACGTTTGAGAAATGGTACTTCAGAAATACATATCAGAACCTGTTGTCATCTTCTTCAT 858
 QY 214 IleIleIleThrMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSer 233
 DB 859 ATTATTATCACCATGACAGAGGTTTGTATCCAGTTTACGTCACCCCTAAGGTGTGATCT 918
 QY 234 AlaPheLeuSerGlyValThrLeuMetLeuLeuThrCysIleValTrpLeuLysLeuVal 253
 DB 919 GCATTTTATCAGGTGTCATCTTGTGTCCTCCTCCTGTCATGTTGTGGCTTAAAGTTGTT 978
 QY 254 SerTyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsn 273
 DB 979 TCTTATGCTCATACTAGCTATGATACATAAGATCCCTAGCCCAATCGAGCTGATAGGCCAAT 1038
 QY 274 ProGluValSerTyrTyrValSerLeuLysSerLeuAlaTyrPheMetValAlaProThr 293
 DB 1039 CCTGAAGTCTCTACTACTAGTTAGCTTGAAGAGCTTGGCATATTCATGCTGCTGCCACA 1098
 QY 294 LeuCystTyrGlnProSerTyrProArgSerAlaCysIleArgLysGlyTrpValAlaArg 313
 DB 1099 TTGTGTTATCAGCAAGTTATCCAGCTTCGTCATGATATACGGAAGGTTGGTGGCTCGT 1158
 QY 314 GlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleLeuGlnTyrIle 333
 DB 1159 CAATTTGCAAACTGGTCATATTCACCGGATTTCACGGATTATATAGAACAAATATATA 1218
 QY 334 AsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeuLeuIleGlu 353
 DB 1219 AATCCCTATTGTCAGGAACCTCAAGACATCTTTGAAAGGCGATCTCTATATGCTATTGAA 1278
 QY 354 ArgValLeuLysLeuSerValProAsnLeuTyrValTrpLeuCysMetPheTyrCysPhe 373
 DB 1279 AGAGTGTGAGCTTTCAGTTCCAAATTTATATGTGTGGCTCTGCAATGTTCTACTGCTTC 1338
 QY 374 PheHisLeuTrpLeuAsnIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyr 393
 DB 1339 TTCCACCTTTGGTTAAACATATTTGGCAGACCTTCTGCTTCGGGGATCGTGAATCTTAC 1398
 QY 394 LysAspTrpTrpAsnAlaLysSerValGlyAspTyrTrpArgMetTrpAsnMetProVal 413
 DB 1399 AAAGATTGTGGAATGCAAAAGTGTGGAGATTACTGGAGAAATGTGGAATATGCTGCTGT 1458
 QY 414 HisLysTrpMetValArgHisIleTyrPheProCysLeuArgSerLysIleProLysThr 433

XX

In the present invention, cDNA libraries from Arabidopsis, corn, rice, soybean and wheat were screened for sequences with homology to a putative acyl CoA cholesterol acyltransferase related gene from Arabidopsis thaliana and diacylglycerol acyltransferases from Homo sapiens and Mus musculus. The cDNA clones identified from this process were used to form complete diacylglycerol acyltransferase cDNA sequences. The present sequence is Arabidopsis thaliana diacylglycerol acyltransferase cDNA. Diacylglycerol acyltransferases are involved in the synthesis of triacylglycerols. Alteration of the expression of the diacylglycerol acyltransferase DNA can be useful for increasing the level of oils in plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful as herbicides.

XX Sequence 1888 BP; 457 A; 394 C; 429 G; 608 T; 0 other;

Alignment Scores:

Pred. No.: 2,19e-252 Length: 1888
Score: 2753.00 Matches: 518
Percent Similarity: 99.62% Conservative: 0
Best Local Similarity: 99.62% Mismatches: 2
Query Match: 99.35% Indels: 0
DB: 21 Gaps: 0

US-09-623-514a-2 (1-520) x AAA48932 (1-1888)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrValThrValThrGluAsnGlyGlyGluPhe 20
DB 131 ATGGCGATTTTGGATTCCTGCTGGCGTTACTACGGTGCAGGAGAACGGTGGCGGAGATTC 190
QY 21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40
DB 191 GTCGATCTTGTAGAGCTGCTGCAGCGAAATCGAGATCGGATCTTCTAACGGACTTC 250
QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60
DB 251 CTCCTCGGTTCGATAAATCTCTCCGGATGATGTGGAGCTCCCGCCAGCTTAGG 310
QY 61 AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
DB 311 GATCGGATTTGATTCGTTTAAACGATGACGCTCAGGGAACACCAATTTGGCCGAGAT 370
QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGly 100
DB 371 AATAACGGTGGTGGCGATAAATACGGTGGTGGGAAGCGCGGAGGAAGGAAGGAAC 430
QY 101 AlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgArgAlaArgGluSer 120
DB 431 GCCGATGTACGTTTACGTATCGACCGCGGTTCAGCTCATCGGAGGCGGAGAGAGT 490
QY 121 ProLeuSerSerAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140
DB 491 CCACCTTAGCTCCGCGCAATCTTCAACAGACGCCATGCCGGATTATTCACCTCTGTGTA 550
QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTrp 160
DB 551 GTAGTCTTATTGCTGTAACAGTAGACTCATCATCGAAATCTTATGAAGTATGGTTGG 610
QY 161 LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180
DB 611 TTGATCAGAACGGATTTCTGGTTTATGTTCAAGATCGTCGAGATTTGGCCGCTTTTCATG 670
QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200
DB 671 TGTGGATATCCCTTTCGATCTTCTTGGCTGCTTTACGGTTGAGAAATGGTACTT 730
QY 201 GlnLysTyrIleSerGluProValIlePheLeuHisIleIleIleThrMetThrGlu 220
DB 731 CAGAAATACATATCAGAACCTGTTGGCATCTTCTTCATATTTATATCACCATGACAGAG 790
QY 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
DB 791 GTTTGTATCCAGTTTACGTACCCCTAAGGTGTGATCTGCTTTTATATCAGGTGTCACT 850

QY 241 LeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTyrAlaHisThrSerTyr 260
DB 851 TTGATGCTCCTCACITTCGATTTGGTGGCTAAAGTTGGTTTCTTATGCTCATCTACTAT 910
QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTrpVal 280
DB 911 GACATAAGATCCCTAGCCAAATGAGCTGATAAGGCCAATCCTGAAGTCTCTACTACGTT 970
QY 281 SerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyr 300
DB 971 ACCTTGAAGAGCTTGGCATATTTTCATGGTGGCTCCCAATGTTGTTATCAGCAAGTTAT 1030
QY 301 ProArgSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIle 320
DB 1031 CCACGTTCTGCTATATACGAAGGGTGGTGGCTCGTCAATTTGCCAAACTGGTTCATA 1090
QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSer 340
DB 1091 TTCACCGGATTCATGGGATTTATATAGAACAAATATATAATCTATTGTCAGGAATCA 1150
QY 341 LysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerVal 360
DB 1151 AAGCATCTCTTGAAGGGCGATCTCTATATGATCTTGAAGAGTGTGAAGCTTTCAGTT 1210
QY 361 ProAsnLeuTyrValTrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIle 380
DB 1211 CCAAAATTTATATGTTGGCTCTGCATGTTCTACTGCTTCTCCACCTTTGGTAAACATA 1270
QY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLys 400
DB 1271 TTGGCAGAGCTTCTCTGCTTCGGGATCGTGAATCTACAAAGATTTGGTGAATGCNAA 1330
QY 401 SerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
DB 1331 AGTGTGGGAGATTACTGGAGATGTGAATATGCCCTGTTTCATAAATGGATGGTTCGACAT 1390
QY 421 IleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIleAlaPhe 440
DB 1391 ATATACTTCCCGTCTTGGCAGCAAGATACCAAGACACTGCCATTATCATTTGCTTTC 1450
QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460
DB 1451 CTAGTCTCTGCACTCTTTCATGAGTATGATCATCGCAGTTCCTTGTGCTCTCTCAAGCTA 1510
QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln 480
DB 1511 TGGGCTTTTCTTGGGATTAATGTTTTCAGGTGCCCTTTGGTCTTCTCATCAAACTATCTACAG 1570
QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500
DB 1571 GAAAGTTTGGCTCAACGGTGGGGAACATGATCTTCTGGTTCATCTTCTGCTTTCGGA 1630
QY 501 GlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520
DB 1631 CAACCGATGTGTGCTTCTTTATTACCACGACCTGATGAACGGAAGGATCGATGTCA 1690
RESULT 6
AAA51484
ID AAA51484 standard; cDNA; 1985 BP.
XX
AC AAA51484;
XX
DT 09-OCT-2000 (first entry)
XX
DE A. thaliana ASL diacylglycerol acyltransferase cDNA.
XX
KW DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW size; weight; carbon flux; TAG1; insertion mutant; ss.
XX
OS Arabidopsis thaliana.
XX
FH Key Location/Qualifiers

QY 1 MetAlaIleuAspSerAlaGlyValThrThrValThrGluAsnGlyGlyGluPhe 20
 DB 237 ATGGCGATTGGAATCTCTGCGGTACTACGGTGACGAGAACGGTGGCGAGAGTTC 296
 QY 21 ValAspLeuAspArgLeuArgArgArgSerArgSerAspSerSerAsnGlyLeuLeu 40
 DB 297 GTGCGATCTTGATAGGCTTCTGTCAGCGAATCGAGATCGGATCTTCTTACGGACCTCTT 356
 QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60
 DB 357 CTCTCTGGTTCGATATAATCTCTCGGATGATGTTGGAGCTCCCGCGAGCTTAGG 416
 QY 61 AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
 DB 417 GATCGGATGATGATTCCTGTTTACGATGACGCTCAGGGAACGCCAATTTGGCGGAGAT 476
 QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGluGlyArgGlyAsn 100
 DB 477 AATAAGGTGGTGGCGATATAACGGTGGTGGNAGAGCGCGGAGAGGAGGAAAC 536
 QY 101 AlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgAlaArgGluSer 120
 DB 537 GCGGATGCTACGTTTACGATACGCGTCCGCTCCAGCTCATCGGAGGCGGAGAGAGT 596
 QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140
 DB 597 CCACCTAGTCTCGACCAATCTCAACAGAGCCATGCCGGAATATTCAACCTCTCTGTA 656
 QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTrp 160
 DB 657 GTAGTCTTATTGCTGTAACAGTAGACTCATCATCGAANAATCTTATGAAGTATGTTGG 716
 QY 161 LeuIleArgThrAspPheThrPheSerArgSerArgSerLeuArgAspTrpProLeuPheMet 180
 DB 717 TTGATCAGAACGGATTTCTGTTTGTAGTTCAAGATCGCTCGAGATTTGGCGCTTTTCATG 776
 QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200
 DB 777 TGTGTATATCCCTTTCGATCTTCTTGGTGGCTTACGGTTCAGAAATTTGGTACTT 836
 QY 201 GlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220
 DB 837 CAGAAATACATATCAGAACCTGTTGTCATCTTCTTCATATATATATACCATGACAGAG 896
 QY 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
 DB 897 GTTTTGTATCCAGTTTACGTCACCTTAAGGTGTGATCTGCTTTTATCAGGTGTCACT 956
 QY 241 LeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTyrAlaHisThrSerTyr 260
 DB 957 TTGATGCTCCTCACCTTGCATTTGTTGGCTAAAGTTGGTTTCTTATGCTCATACTAGTAT 1016
 QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTyrVal 280
 DB 1017 GACATAAGATCCCTAGCCAAATGACGTGATAGGCCAATCTTGAAGTCTCTACTACTGTT 1076
 QY 281 SerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyr 300
 DB 1077 AGCTTGAAGAGCTTGGCATATTTCATGTCGCTCCACATTTGTTATCAGCCAAAGTTAT 1136
 QY 301 ProArgSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIle 320
 DB 1137 CCACGTTCTGCGATGATATACGAAGGGTGGGTGGCTCGCAATTTGCAAACTGGTCAATA 1196
 QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSer 340
 DB 1197 TTCACGGGATTCATGGATTTATATAGAACATATATAAATCTTATGTCAGGAAGTCA 1256
 QY 341 LysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerVal 360
 DB 1257 AAGCATCCCTTGAAGGGCATCTTCTATATGCTATTGAAAGAGTGTGAAGCTTTTCAGTT 1316

QY 361 ProAsnLeuTyrValTrpTrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIle 380
 DB 1317 CCAAAATTTATATGTGTGGCTGTCATGTTCTACTGTTCTTCCACCTTTGGTTAAACATA 1376
 QY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLys 400
 DB 1377 TTGGCAGAGCTTCTCTGCTTCGGGATCGTGAATTTCTACAAAGATTTGGTGAATGCAAAA 1436
 QY 401 SerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
 DB 1437 AGTGTGGAGATTACTTGGAGAAATGCGAATATGCCGTGTTCAATAATGGATGGTTCGACAT 1496
 QY 421 IleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPhe 440
 DB 1497 ATATAGTTCCTCGGCTTGGCAGCAAGATACCAAGACACTCCCAATATCATTTGCTTTC 1556
 QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460
 DB 1557 CTAGTCTCTGCGAGTCTTTTCATGAGCTATGCTATGCGAGTTCCTTGTCTCTCTCAAGCTA 1616
 QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln 480
 DB 1617 TGGGCTTTTCTTGGGATTTATGTTTCAGGTGCTTGTCTTTCATCAAACTATCTACAG 1676
 QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500
 DB 1677 GAAAGTTTGGCTCAACGGTGGGGAACATGATCTTCTGTTTCATCTCTGCAATTTTCGGA 1736
 QY 501 GlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520
 DB 1737 CAACCGATGTGTGCTTCTTTATTACCGAGCTCATGACCGAAGGATCGATGTCA 1796
 RESULT 5
 -AAA48932
 ID AAA48932 standard; cDNA; 1888 BP.
 XX
 AC AAA48932;
 XX
 DT 06-DEC-2000 (first entry)
 XX
 DE Arabidopsis diacylglycerol acyltransferase cDNA.
 XX
 KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
 KW triacylglycerol; herbicide; EC2.3.1.20; SS.
 XX
 OS Arabidopsis thaliana.
 XX
 FH Key Location/Qualifiers
 CDS 131..1693
 FT /tag- a
 FT /product= Diacylglycerol_acyltransferase
 XX
 PN W0200032756-A2.
 XX
 PD 08-JUN-2000.
 XX
 PF 01-DEC-1999; 99WO-US28354.
 XX
 PR 02-DEC-1998; 98US-0110602.
 PR 31-MAR-1999; 99US-0127111.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Cahoon EB, Kinney AJ, Cahoon RE;
 XX
 DR WPI: 2000-412308/35.
 DR P-PSDB; AAY94512.
 XX
 PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for
 PT synthesis of triacylglycerols and increasing the level of oils in plant
 PT seeds
 XX
 PS Claim 4; Page 38; 62pp; English.

```

Db 717 TTGATCAGACGGATTTCGTTAGTTCAAGATCGCTGCAGATTGCCCCCTTTTCATG 776
Qy 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200
Db 777 TGGTTGATATCCCTTTCGATCTTCCCTTTGGCTGCCCTTACGGTTGAGAAATGGTACTT 836
Qy 201 GlnLysTrpIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220
Db 837 CAGAAATACATATCAGAACCTGTTGTCATCTTCTTATATATATATATATATATATATAT 896
Qy 221 ValLeuTrpProValTrpValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
Db 897 GTTTTGATACCGTTTACCTCACCTAAGGTGTCATCTGCTTTTATCAGGTGTCAT 956
Qy 241 LeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTrpAlaHisThrSerTrp 260
Db 957 TTGATGCTCCCTCACTTGCTATGTTGGCTAAAGTTGGTTTCTTATGCTCATCTAGCTAT 1016
Qy 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTrpVal 280
Db 1017 GACATAAGATCCCTAGGCCATGCACTGATAGGCCAATCCTGAAGTCTCTCTACTAGCTT 1076
Qy 281 SerLeuLysSerLeuAlaTrpPheMetValAlaProThrLeuCysTrpGlnProSerTrp 300
Db 1077 AGCTTGAAAGCTTGGCATATTTTCATGCTGCTCCACATTTGTTTATCAGCCAAAGTTAT 1136
Qy 301 ProArgSerAlaCysIleArgLysGlyTrpValAlaAlaArgGlnPheAlaLysLeuValIle 320
Db 1137 CCACGTTCTGCATGATACGGAAGGTTGGTGGCTGCTCAATTTGCNAAACTGGTCATA 1196
Qy 321 PheThrGlyPheMetGlyPheIleIleGluGlnTrpIleAsnProIleValArgAsnSer 340
Db 1197 TTCACCGATTTCATGGGATTATATAGAACAAATATATATATATATATATATATATATAT 1256
Qy 341 LysHisProLeuLysGlyAspLeuLeuTrpAlaIleGluArgValLeuLysLeuSerVal 360
Db 1257 AAGCATCTTTGAAAGCGATCTTCTATATGCTATTGAAAGAGTGTGGAAGCTTTTCAGTT 1316
Qy 361 ProAsnLeuTrpValTrpLeuCysMetPheTrpCysPhePheHisLeuTrpLeuAsnIle 380
Db 1317 CCAAAATTATATGTTGGCTGTCATGTTCTACTGCTTCTCCACCTTTGGTTAAACATA 1376
Qy 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTrpLysAspTrpTrpAsnAlaLys 400
Db 1377 TTGCAGAGCTTCTCTGCTCGGGGATCGTGAATCTACAAAGATTGGTGAATGCAGAAA 1436
Qy 401 SerValGlyAspTrpTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
Db 1437 AGTGTGGAGATTACTGGAGATTGGAAATATGCTGTTCAATGATGGATGGTTTCGACAT 1496
Qy 421 IleTrpPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPhe 440
Db 1497 ATATACCTCCGCTGTTGGCAGCAAGATACCAAGACACTCGCCATATCATTTGCTTTC 1556
Qy 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460
Db 1557 CTAGTCTCTGCAGTCTTTCATGAGCTATGATCGAGTTCCTGCTGCTCTCTCAAGCTA 1616
Qy 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTrpLeuGln 480
Db 1617 TGGGCTTTCTTGGGATATGTTTTCAGGTGCTTTGGTCTTCATCAACAAACTATCTACAG 1676
Qy 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500
Db 1677 GAAAGTTTGGCTCAACGGTGGGGAACATGATCTCTGTTTCATCTCTCTGCTATTTCCGA 1736
Qy 501 GlnProMetCysValLeuLeuTrpTrpHisAspLeuMetAsnArgLysGlySerMetSer 520
Db 1737 CAACCGATGTGTGTCTTCTTATATACCAAGCTGATGAACCGAAAGGATGATGATCA 1796

```

RESULT 4
AAS01106

ID AAS01106 standard; cDNA; 1942 BP.

```

XX AC AAS01106;
XX 31-MAY-2001 (first entry)
XX Arabidopsis thaliana sterol acyltransferase ACAT EST sequence.
XX Lecithin:cholesterol acyltransferase-like; LCAT; sterol acyltransferase;
XX acyl CoA:cholesterol acyltransferase-like; ACAT; plant oil; cooking oil;
XX nutritional supplement; dairy product; food product; salad dressing;
XX expressed sequence tag; EST; ss.
XX Arabidopsis thaliana.
XX Key Location/Qualifiers
XX CDS 237..1799
XX /*tag= a
XX /product= "ACAT"
XX WO200116308-A2.
XX 08-MAR-2001.
XX 30-AUG-2000; 2000MO-US23863.
XX 30-AUG-1999; 99US-0152493.
XX (MONS ) MONSANTO CO.
XX Lassner M, Van Eenennaam A;
XX WPI; 2001-169010/17.
XX P-PSDB; AAU00462.

```

New isolated nucleic acid encoding plant lecithin:cholesterol acyltransferase-like or acyl (coenzyme A) CoA:cholesterol acyltransferase-like polypeptides, for modifying the sterol content and oil production of plants -

Claim 4; Page 105-106; 127pp; English.

The present sequence encodes for Arabidopsis thaliana acyl CoA:cholesterol acyltransferase-like (ACAT). Several novel polynucleotides encoding the plant sterol acyltransferases LCAT (lecithin:cholesterol acyltransferase-like; AAS001081-AAS01104, AAS01341) and ACAT (AAS01311-AAS01319) are described. A yeast LCAT related open reading frame, LRO1 gene sequence (AAS01342), and a rat ACAT (AAS01105) cDNA sequence are also described. The polynucleotides encoding LCAT or ACAT are used to produce LCAT or ACAT polypeptides. They can also be used in a recombinant construct to transform a host cell (preferably of a plant) or a plant. The recombinant construct is used to increase or decrease the sterol content of the host cell or plant. It can be used to alter oil production of the cell or plant, preferably by increasing it. The oil of the plant or the plant itself is used as a food product, or as nutritional or dietary supplements, or in pharmaceutical compositions for lowering cholesterol. The oil can be used in foods e.g. margarine, butter, cooking oil, and dressings e.g. salad dressings, mayonnaise, cheese, processed meat, pasta, sauces, cereals, desserts, dips, chips, baked goods, pastries, cookies, snack bars, confections, chocolates, and beverages. The alteration in sterol content and/or composition can also provide a plant with tolerance to stress and insect damage.

Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

Alignment Scores:

```

Pred. No.: 4.39e-254 Length: 1942
Score: 2771.00 Matches: 520
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Db: 22 Gaps: 0

```

US-09-623-514A-2 (1-520) x AAS01106 (1-1942)

QY 361 ProAsnLeuTyrValTrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIle 380
 DB 1317 CCAAAATTTATATGCTGGCTGCAATGTTCTACTCTTCCACCTTTGGTTAAACATA 1376
 QY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpAsnAlaLys 400
 DB 1377 TGGCAGACCTCTCTGCTCGGGATCGTGAATCTCAAAAGATTGGTGAATGCANAA 1436
 QY 401 SerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
 DB 1437 AGTGTGGAGATCTACTGGAGATGTGGAATATGCTGTTTCATAAATGATGGTTCACAT 1496
 QY 421 IleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPhe 440
 DB 1497 ATATACTTCCCTGCTGTCGGCAGACAGATACCAAGACACTCGCCATTATCATGTGCTTC 1556
 QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460
 DB 1557 CTAGCTCTGCACTCTTCATGAGCTATGCAATGCGAGTCTCTGCTCTTCAAGCTA 1616
 QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln 480
 DB 1617 TGGGCTTTCTGGGATTATGTTTCAGGTGCTTGTCTTCATCACAACATATCTACAG 1676
 QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500
 DB 1677 GAAAGGTTTGGTCAACGGTGGGACATGATCTTCTGTTTCATCTCTGCAATTTTCGGA 1736
 QY 501 GlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520
 DB 1737 CAACCGATGCTGCTGCTCTTTATACACGACCTGATGACCGAAGATCGATGTCA 1796
 RESULT 3
 AAZ45371
 ID AAZ45371 standard; DNA; 1942 BP.
 XX
 AC AAZ45371;
 XX
 DT 27-MAR-2000 (first entry)
 XX
 DE Acyl-CoA:cholesterol acyltransferase (ACAT)-like protein DNA.
 XX
 KW Acyl-CoA:cholesterol acyltransferase; ACAT; ACAT-like protein; sterol;
 KW ester; triacylglycerol; fatty acyl-CoA; lipid composition; plant cell;
 KW diacylglycerol acyltransferase; DAGAT; triglyceride; cancer; diabetes;
 KW cardiopulmonary disease; heart failure; atherosclerosis; adipocytosis;
 KW leukaemia; skin carcinoma; fibroblastoma; metabolic disorder; obesity;
 KW abnormal lipid metabolism; abnormal fat absorption;
 KW lipoprotein secretion; adipogenesis; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 FH Key Location/Qualifiers
 FT CDS 237..1799
 FT /tag- a
 FT /product- "Acyl-CoA:cholesterol acyltransferase (ACAT)-
 FT like protein"
 XX
 PN W09963096-A2.
 XX
 PD 09-DEC-1999.
 XX
 PF 04-JUN-1999; 99WO-US12541.
 XX
 PR 05-JUN-1998; 98US-0088143.
 PR 12-NOV-1998; 98US-0108389.
 XX
 PA (CALJ) CALGENE LLC.
 XX
 PI Lassner MW, Ruezinsky DW;
 XX
 DR WPI; 2000-105701/09.

DR P-PDB; AAY54143.

XX Novel polynucleotides used for modifying plant oil composition and for
 PT developing products for treating e.g. cancer, diabetes, cardiopulmonary
 PT disease or metabolic disorders

XX Claim 4; Fig 1; 89pp; English.

XX The present sequence encodes an acyl-CoA:cholesterol acyltransferase
 CC (ACAT) related protein. The ACAT-like protein is active in the formation
 CC of a sterol, ester and/or triacylglycerol from a fatty acyl-CoA and
 CC sterol and/or diacylglycerol substrate. The DNA can be used for
 CC modifying the lipid composition of plant cells. The ACAT-like protein
 CC has diacylglycerol acyltransferase (DAGAT) activity, and so the
 CC synthesis of triglycerides can be suppressed or increased using the
 CC DNA. The protein can be used to produce plant oils with a modified
 CC triglyceride content. The products can also be used to identify
 CC antagonists and agonists of DAGAT activity. Such agonists and
 CC antagonists are particularly useful in treating or ameliorating
 CC diseases associated with DAGAT activity, including diseases associated
 CC with altered cellular diacylglycerol concentration or PKC activity,
 CC including cancer, diabetes, cardiopulmonary diseases e.g. heart failure,
 CC atherosclerosis, adipocytosis, leukaemia, skin carcinoma, fibroblastoma,
 CC metabolic disorders, obesity, diseases associated with abnormal lipid
 CC metabolism, and diseases associated with abnormal fat absorption,
 CC lipoprotein secretion and adipogenesis.

XX Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

Alignment Scores:

Pred. No.: 4,39e-254 Length: 1942
 Score: 2771.00 Matches: 520
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-623-514A-2 (1-520) x AAZ45371 (1-1942)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGlyGluPhe 20
 DB 237 ATGGCGATTGTGGATTCTGCTGGCGTTACTACGGTCACGGAGAACGGTGGCGGAGTTC 296
 QY 21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40
 DB 297 GTCGATCTTGATAGGCTTCGTCGACGAAATCGAGATCGGATCTTCTTCAACGACTCTT 356
 QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyValProAlaAspValArg 60
 DB 357 CTCTCTGGTTCGATTAATATCTCTCTCGGATGATGTTGGAGCTCCCGCCGACGTAGG 416
 QY 61 AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
 DB 417 GATCGGATTGATTCGTTTAAAGATGACGCTCAGGAAACAGCCAAATTTGGCCGGAGAT 476
 QY 81 AsnAsnGlyGlyCysAspAsnAsnGlyGlyArgGlyGlyGlyGlyGlyArgGlyAsn 100
 DB 477 AATAACGGTGGTGGCGATATAACCGTGTGGAAAGAGCGCGGAGAGAGAGAAAC 536
 QY 101 AlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgAlaArgGluSer 120
 DB 537 GCCGATGCTACGTTTACGTTATCGACGCTCGGTTCAGCTCATCGGAGGCGGAGAGAGT 596
 QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140
 DB 597 CCACATTAGCTCCGACGCAATCTTCAACAGAGCCATGCGCGGATTTATCAACCTCTGTGA 656
 QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTrp 160
 DB 657 GTAGTTCTTATTGCTGTAACAGTAGTACATCATCGAAAAATCTTATGAAGTATGTTGG 716
 QY 161 LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180

QY 501 GlnPrometCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520
 AAA88835
 Db 1639 CAACCGATGTGTGCTTCTTTATACACGACCTGATGAACCGAAAGGATCGATGTCA 1698

RESULT 2
 ID AAA88835 standard; DNA; 1942 BP.
 XX
 AC AAA88835;
 XX
 DT 19-FEB-2001 (first entry)
 XX
 DE Arabidopsis acyl CoA:cholesterol acyltransferase DNA.
 XX
 KW Acyl CoA:cholesterol acyltransferase; ACAT; transgenic plant;
 KW sterol; tocopherol; phytosterol; phytostanol; anticholesterolemic;
 KW hypolipemic; ss.
 XX
 OS Arabidopsis thaliana.
 XX
 FH Key Location/Qualifiers
 FT CDS 237..1799
 FT /*tag= a
 FT /*EC_number= "2.3.1.26"
 XX
 PW WO2000061771-A2.
 XX
 PD 19-OCT-2000.
 XX
 PF 12-APR-2000; 2000WO-US09696.
 XX
 PR 12-APR-1999; 99US-0128995.
 XX
 PA (MONS) MONSANTO CO.
 XX
 PI Venkatramesh M, Corbin DR, Bhat GB, Boddupalli SS, Grebenok RJ;
 PI Kishore GM, Lardizabal KD, Lassner MW, Rangwala SH, Karunanandaa B;
 XX
 DR WP1; 2000-665136/64.
 DR P-PSDB; AAB19740.
 XX
 PT Genetically engineering the biosynthetic pathways in plants involved in
 PT the accumulation of sterol compounds and tocopherol to produce
 PT compounds for lowering the level of low density lipoprotein cholesterol
 PT in blood serum -
 XX
 PS Disclosure; Page 57-58; 166pp; English.
 XX
 CC The present sequence is that of Arabidopsis thaliana DNA encoding
 CC acyl CoA:cholesterol acyltransferase (ACAT, see AAB19740). Sterol
 CC O-acyltransferases such as ACAT catalyse the formation of
 CC cholesterol esters from cholesterol and long chain fatty acids.
 CC Recombinant constructs of the invention are used to alter the
 CC biosynthesis and accumulation of sterols and tocopherols in
 CC transgenic plants. Seeds of such plants may contain elevated
 CC levels of sitosterol and/or its esters, and alpha-tocopherol, and
 CC reduced levels of campesterol and campestanol and their esters.
 CC The seeds may also contain the novel sterol brassicatanol. Oil
 CC obtained from the seeds can be used in food and pharmaceutical
 CC compositions to lower levels of low density lipoprotein cholesterol
 CC in blood serum. ACAT enzymes can be used in the present invention
 CC to produce elevated levels of phytosterol and/or phytostanol esters.
 XX
 SQ Sequence 1942 BP; 461 A; 421 C; 425 G; 635 T; 0 other;

Alignment Scores:
 Pred. No.: 4,39e-254 Length: 1942
 Score: 2771.00 Matches: 520
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 21 Gaps: 0

US-09-623-514a-2 (1-520) x AAA88835 (1-1942)

QY 1 MetalalleLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGlyGluPhe 20
 Db ATGGCGATTTGGATTCTGCTGGCGTTACTACGGTGACGGAGACGGTGGCGGAGGTTTC 296
 QY 21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40
 Db GTCGATCTTGATAGGCTTCGTCGACGGAATCGAGATCGGATTCCTCTAACGGACTCTT 356
 QY 41 LeuSerGlySerAspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArg 60
 Db CTCCTCGTTCCGATAAATTCCTTCGGATGATGTGGAGCTCCGCCGACCTTAGG 416
 QY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
 Db GATCGGATTGATTCGGTTGTTAACGATCAGCTCAGGGAACAGCCAAATTTGCCGGAGAT 476
 QY 81 AsnAsnGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGly 100
 Db AATAACGGTGGTGGCGATAAATAACGGTGGTGAAGAGCGCGGAGAGGAAGAGGAAC 536
 QY 101 AlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgArgAlaArgGluSer 120
 Db GCCGATGCTACGTTTACGATATCGACCGTTCGGTTCAGCTCATCGGAGGGCGGAGAGAGT 596
 QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140
 Db CCATTAGCTCGGCGCAATCTTCAACAGAGCCATCGCGGATATTATCACTCTGTGTGA 656
 QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTrp 160
 Db GTAGTTCTTATTGCTGTAAACAGTAGACTCATCATCGAAATCTTATCAAGTATGTTGG 716
 QY 161 LeuIleArgThrAspPheTrpPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180
 Db TTGATCAGAACGGATTTCTGTTTAAAGATCGCTCGGAGATTTGCCGCTTTTCATG 776
 QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200
 Db TGTGTATATCCCTTTCGATCTTCTTGGTGGCTTTACGGTTGAGAAATTTGTACTT 836
 QY 201 GlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleIleThrMetThrGlu 220
 Db CAGAAATACATATCAGAACCTGTTGTCATCTTCTTCATATATTATATCACCATGACAG 896
 QY 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
 Db GTTTTGTATCCAGTTTACGTCACCTTAAGGTGTGATTCCTTTTATCAGGTGTCAC 956
 QY 241 LeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTyrAlaHisThrSerTyr 260
 Db TTGATGTCCTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1016
 QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTrpVal 280
 Db GACATAAGATCCCTAGCCAATGCAGCTGATAGGCCAATCCTGAAGTCTCTCTACTACCT 1076
 QY 281 SerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyr 300
 Db AGCTTGAAGAGCTTGGCATATTTTCATGTCGCTCCCATATTTGCTTATCAGCAAGTTAT 1136
 QY 301 ProArgSerAlaCysIleArgLysGlyTrpValAlaArgGlnPheAlaLysLeuValIle 320
 Db CCAGTCTGCTGATGATACGGAAGGTTGGTGGCTGCTCAATTTGCAAACTGGTCATA 1196
 QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSer 340
 Db TTCACCGGATTCATGGGATTTATAATAGAACAAATATAAATCTTATTTGTCAGGAAC 1256
 QY 341 LysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerVal 360
 Db AAGCATCCTTTTGAAGGCGATCTTCTATATGCTATTTGAAGAGAGTGTGTAAGCTTT 1316

/*tag- a
/product= Diacylglycerol_acyltransferase

W0200036114-A1.

22-JUN-2000.

16-DEC-1999; 99WO-CA01202.

17-DEC-1998; 98US-0112812.

(CANADA) NAT RES COUNCIL CANADA.

Zou J, Taylor DC, Wei Y, Jako CC;

WPI; 2000-431592/37.

P-PSDB; AAY96853.

New DNA encoding diacylglycerol acyltransferase from Arabidopsis thaliana for transforming plants and regulating seed oil content, fatty acid synthesis and seed oil acyl composition in commercial and crop plants

Claim 1; Page 64-65; 91pp; English.

This cDNA encodes Arabidopsis thaliana diacylglycerol acyltransferase (DGAT). The DGAT and an insertion mutant (AS11) are useful for regulating seed oil content, the ratio of diacylglycerol/triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil acyl composition, seed size/weight and carbon flux into other seed components in commercial and crop plants. The natural formation of triacylglycerols can be modified to increase the yield in commercial plant oils or modify their composition to achieve specific commercial improvements of plants and plant products.

Sequence 1904 BP; 458 A; 399 C; 431 G; 616 T; 0 other;

Alignment Scores:

Pred. No.: 4,26e-254 Length: 1904
Score: 2771.00 Matches: 520
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 21 Gaps: 0

US-09-623-514A-2 (1-520) x AAA51482 (1-1904)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGlyGluPhe 20
|||||
Db 139 ATGGCGATTGGATTCTGCTGGCGTTACTACGGTCACGGAGAACCGTGGCGGAGAGTTC 198
QY 21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerAsnGlyLeuLeu 40
|||||
Db 199 GTCGATCTTGATAGGCTTCGTCACGCGAAATCGAGATCGGATCTCTTAACGGACTCTT 258
QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60
|||||
Db 259 CTTCTGGTTCGGATAAATCTCTCGGATGATGTTGGAGCTCCCGCCGACGTTAGG 318
QY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
|||||
Db 319 GATCGGATTGATTCGCTGTTAACGATGACGTCACGGAGAACCGCAATTTGGCCGAGAT 378
QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyArgGlyAsn 100
|||||
Db 379 AATAACGGTGGTGGCGATAATAACGGTGGTGGAGAGCGCGGAGAGGAGGAGGAGGAGG 438
QY 101 AlaAspAlaThrPheThrThrArgProSerValProAlaHisArgAlaArgGluSer 120
|||||
Db 439 GCGGATGCTACGTTTACGATACCGCTCGGTTCCAGCTCATCGGAGGCGGAGAGAGT 498
QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140
|||||
Db 499 CCACCTAGTCCGACCGCAATCTCAACAGAGCGCATGCCGGATTATCAACCTCTCTGTGA 558

QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTrpGlyTrp 160
|||||
Db 559 GTAGTTCTTATTGCTGTAAACAGTAGACTCATCATCGAAATCTTATGAAGTATGCTGG 618
QY 161 LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180
|||||
Db 619 TTGATCAGAACGGATTCTGGTTTAGTTTCAAGATCGCTGCGAGATTGGCGCTTTTCATG 678
QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200
|||||
Db 679 TGTGTATATCCCTTTCGATCTTTCTTGGCTGGCTTTACGGTTGAGAAATTTGGTACTT 738
QY 201 GlnLysTrpIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220
|||||
Db 739 CAGAAATACATATCAGAACCTGTTGTCTCTTCTTTCATATATTATTCACCATGACAGAG 798
QY 221 ValLeuTrpProValTrpValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
|||||
Db 799 GTTTTGTATCCAGTTTACGTCACCTAAGGTGTGATTCTGCTTTTATCAGGTGTCAC 858
QY 241 LeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTrpAlaHisThrSerTrp 260
|||||
Db 859 TTGATGCTCTCTCACTTGCAATGTCGTGCTAAAGTTGGTTCTTATGCTCATAGCTAT 918
QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTrpTrpVal 280
|||||
Db 919 GACATAAGATCCCTGACCAATGAGCTGATGAAGGCCAATCTCTGAAGTCTCTACTACGTT 978
QY 281 SerLeuLysSerLeuAlaTrpPheMetValAlaProThrLeuLeuCysTrpGlnProSerTrp 300
|||||
Db 979 AGCTTGAAGAGCTTGGCATAATTTTCATGGTGGCTCCACCATTTGTTATCAGCCAAGTTAT 1038
QY 301 ProArgSerAlaCysIleArgLysGlyTrpValAlaAlaArgGlnPheAlaLysLeuValIle 320
|||||
Db 1039 CCACGTTCTGCATGATATACGGAAGGTTGGTGGCTGCTCAATTTGCAAACTGGTCATA 1098
QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTrpIleAsnProIleValArgAsnSer 340
|||||
Db 1099 TTCACCGGATTCATGGGATTTATAATAGAACAATATATAATCTTATGTCAGGAAGTCA 1158
QY 341 LysHisProLeuLysGlyAspLeuLeuTrpAlaIleGluArgValLeuLysLeuSerVal 360
|||||
Db 1159 AAGCATCCCTTTGAAGGCGCATCTCTATATGCTATTTCAAGAGTGTGTAAGCTTTTCAGTT 1218
QY 361 ProAsnLeuTrpValTrpLeuCysMetPheTrpCysPhePheHisLeuTrpLeuAsnIle 380
|||||
Db 1219 CCAAAATTTATATGTGTGGCTCTGCATGTTCTACTGCTTCTTCCACCTTTGGTTAAACATA 1278
QY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTrpLysAspTrpTrpAsnAlaLys 400
|||||
Db 1279 TTGGCAGAGCTTCTCTGCTTGGGGATCGTGAATTTCTACAAGATTGGTGAATGCAAAA 1338
QY 401 SerValGlyAspTrpTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
|||||
Db 1339 AGTGTGGGAGATTACTGGGAGATGCGGAATATGCCCTGTTCAATAAATGGATGGTTCACAT 1398
QY 421 IleTrpPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPhe 440
|||||
Db 1399 ATATACTTCCCGTGGCTGGCAGCAAGATACCAACACACTCCCATTTATCATGCTTTC 1458
QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460
|||||
Db 1459 CTAGTCTCTGCAGCTTTTCATGAGCTATGATCGCAGTTCCTTGTCTCTCTTCAAGCTA 1518
QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTrpLeuGln 480
|||||
Db 1519 TGGGCTTTTCTGGGATTTATGTTTCAGGTGCCCTTTGCTTCTTCATCAAACTATCTACAG 1578
QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGly 500
|||||
Db 1579 GAAAGGTTTGGCTCAACGGTGGGGAACATGATCTTCTGGTTCACTCTTGCATTTTCGGA 1638

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 31, 2003, 04:59:59 ; Search time 377 Seconds
(without alignments)
3723.363 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771

Sequence: 1 MAILDSAGVTVTENGGEF.....QPMCVLLYYHDLNMRKSGMS 520

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09623514/runat_29082003.152159.46/app_query.fasta_1.711
-DB=N_Geneseq_19Jun03 -QPMV=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Diosum62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -PCT=SCORE=100 -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09623514 -CGN_1.1_471 @runat_29082003.152159.46 -NCPU=3
-NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
- 25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	100.0	2771	1904	21	A. thaliana diacyl
2	100.0	2771	1942	21	Arabidopsis acyl C
3	100.0	2771	1942	21	Acyl-CoA:cholester
4	100.0	2771	1942	22	Arabidopsis thalia
5	99.4	2753	1888	21	Arabidopsis diacyl
6	98.2	2747.5	1985	21	A. thaliana AS11 d
7	80.3	2225	5193	21	A. thaliana diacyl
8	80.3	2216.5	5339	21	A. thaliana AS11 d
9	65.7	1821	1942	21	Soybean diacylglyc
10	57.7	1597.5	1587	21	Rice diacylglycer
11	57.6	1596.5	1975	21	Wheat diacylglycer
12	46.2	1279	1559	21	Corn diacylglycer
13	45.5	1260	1281	21	Corn diacylglycer
14	37.1	1028.5	901	21	Corn diacylglycer
15	28.7	794	1976	21	Human ACAT Related
16	27.9	774	978	21	Corn diacylglycer
17	27.9	773.5	1650	21	Mouse diacylglycer
18	27.9	773.5	1766	21	Rat acyl CoA:chole
19	27.9	773.5	1766	21	Acyl-CoA:cholester
20	27.9	773.5	1766	22	Rat sterol acyltra
21	27.6	764.5	1732	24	Bovine DGAT1 CDNA
22	26.7	738.5	1521	19	Human acylcoenzyme
23	25.6	709.5	1895	21	Human acyl CoA:cho
24	25.6	709.5	1895	21	DNA encoding a pro
25	25.2	698	380	24	Arabidopsis thalia
26	23.5	652.5	629	21	EST with homology
27	23.5	652.5	629	21	A. thaliana diacyl
28	21.0	582	7490	23	Drosophila melanog
29	20.7	573.5	993	22	Human CDNA Seq ID
30	20.7	573.5	993	24	Human polynucleoti
31	19.3	535	1122	20	DNA encoding a hum
32	19.0	526.5	983	19	Human acylcoenzyme
33	19.0	526	470	24	Arabidopsis thalia
34	16.6	459	3996	23	Drosophila melanog
35	15.9	440	275	25	Human GDP-mannose
36	14.9	421	774	22	CDNA encoding nove
37	14.9	413.5	3649	19	Human acylcoenzyme
38	14.9	412.5	4011	15	Acetyl coenzyme A:
39	14.7	407.5	452	24	Human ovarian canc
40	14.5	401.5	3650	19	Human acyl-coenzym
41	13.9	384	11771	24	Bovine DGAT1 gene.
42	13.7	378.5	2421	19	Acylcoenzyme A:cho
43	13.6	377	2421	19	Yeast acyl-coenzym
44	13.6	376	542	22	CDNA encoding nove
45	13.5	375	1607	21	Mouse acyl CoA:cho

ALIGNMENTS

RESULT 1

AAA51482

ID AAA51482 standard; cDNA; 1904 BP.

AC AAA51482;

XX 09-OCT-2000 (first entry)

XX A. thaliana diacylglycerol acyltransferase cDNA.

DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
KW size; weight; carbon flux; ss.

XX Arabidopsis thaliana.

XX Key Location/Qualifiers

FT CDS 139..1701

Db 624 GTTGGTCTATCACCCTCCCTGGTGGTCTGGCATCATCTTC 683
QY 250 LeuLysLeuValSerTyrAlaHisThrSertyr-----Aspileargser 264
Db 684 CTCAGCTTTTCTCTACCGGATGTCATCTGTGTGGCCGCCAGGAGGTCAAGGCC 743
QY 265 LeuAlaAsnAlaAlaAspLys-----AlaAsnProGluValSerTyr 278
Db 744 AAGCTGTGTCTGTCAGGAAGAGTCACTGGGGTGTGTCGCCAGAACACTGTAAGCTAT 803
QY 279 -----TyrValSerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCystyr 296
Db 804 CCGGACAACTGACCTACCGAGATCTATTACTTCTTGTCTTCTTGTGTAT 863
QY 297 GlnProSerTyrProArgSerAlaCysIleArgLysGlyTyrPheValAlaArgGlnPheAla 316
Db 864 GAACCTCACTTCTTCGATCCCGCCGAAATACGAAGCGTTTCTGCTACGGGGGGTCTT 923
QY 317 LysLeuValIlePheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIle 336
Db 924 GAGATGCTCTTTTCCACCGAGTCTCAAGTGGGGCTGATCCAGCAGTGGATGTCCTACT 983
QY 337 ValArgAsnSerLysHisProLeuLysGlyAspLeuLeuTyrAla-----IleGluArg 354
Db 984 ATCCAGAACTCCATGAAGCCCTCAAG--GACATGGACTATTCAGAACTCATTTAGCGGT 1040
QY 355 ValLeuLysLeuSerValProAsnLeuTyrValTyrLeuCysMetPheTyrCysPhePhe 374
Db 1041 CTTCTAAAGCTGGCGGTCCCAACCATCTGATATGCTCATCTTCTTATTGGCTTTC 1100
QY 375 HisLeuTyrLeuAsnIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLys 394
Db 1101 CACTCATGCTCAATGCTGTGGCAGAGCTCCTGCAGTTTGGAGACCGGAGTCTACAGG 1160
QY 395 AspTyrTrpAsnAlaLysSerValGlyAspTyrTrpArgMetTrpAsnMetProValHis 414
Db 1161 GACTGGTGAATGCTGAGTGTACCTACTTTTGGCAGAACTGGAATATCCCGTGCAC 1220
QY 415 LysTrpMetValArgHisIleTyrPheProCysLeuArgSerLysIleProLysThrLeu 434
Db 1221 AAGTGTGCATCAGACACTTCTACAAGCTATGCTCAGACTGGGCAGCAACAATGGATG 1280
QY 435 AlaIleIleIleAlaPheLeuValSerAlaValPheHisGluLeuCysIleAlaValPro 454
Db 1281 GCCAGGACTGGGGTCTTTTGGCGTCAGCCTCTTCCATGAGTACCTAGTGAGCATCC 1340
QY 455 CysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuValPhe 474
Db 1341 CTGAGGATGTTCCGGCTCTGGGCATTCACGCCATGATGGCTCAGTCCACTGCCCTGG 1400
QY 475 IleThrAsnTyrLeuGlnGluArgPheGlySerThrValGlyAsnMetIlePheTrpPhe 494
Db 1401 ATTGTGAACCGGCTC-----TTCCAAGGAAGTATGGCAATGGCAGCTGTGTGG--- 1448
QY 495 IlePheCysIlePheGlyGlnProMetCysValLeuLeuTyrTyrHisAsp 511
Db 1449 GTGACACTCATATTGGCAACCGGTGCTGTGCTCATGTATGTCCAGGAC 1499

Search completed: August 31, 2003, 08:27:19

Job time : 403 secs

Qy	225	ValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeu	244
		:	
Db	660	GCAGCTGTGCCCTACTGGTGTAGTCTCATCTCACCTCCAGTGGGTTCGTTGTTGCTCTGC	719
Qy	245	Thr--CysIleValTrpLeuValSerTyrAlaHisThrSerTyrAspIleArg	263
		:	
Db	720	TCATATCCCATCATGTTTCCCTCAAGCTTTATTCCTACCGGGATGTCACAGTGTGGTGGCCG	779
Qy	264	-----SerLeuAlaAsnAlaAlaAsp	270
		:	
Db	780	CAGCGAGGCTCAAGGCCAAGCTGTCTACAGGGAAGAGTCAAGTGGGGGTGCT---	836
Qy	271	LysAlaAsnProGluValSerTyr-----TyrValSerLeuLysSerLeuAlaTyrPhe	288
		:	
Db	837	---GCCCAAGCAGCTGTGAGCTATCCAGACAACTGTACCTACCGAGATCTCTATTACTTC	893
Qy	289	MetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArgLys	308
		:	
Db	894	ATCTTTGCTCTCTACTTTTGTGTTATGAACTCAACTTCTCGTCCGCCCGCAATACGAAG	953
Qy	309	GlyTrpValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIle	328
		:	
Db	954	CGCTTCTGTACGACGAGTCTTGAGATGCTCTTTTACCAGCTTCAAAGTGGGGCTG	1013
Qy	329	IleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeu	348
		:	
Db	1014	ATCCACAGTGGATGGTCCCTACTATCCCAACTCCATGAAGCCCTTCAAG---GATATG	1070
Qy	349	LeuTyrAla-----IleGluArgValLeuLysLeuSerValProAsnLeuTyrValTrp	366
		:	
Db	1071	GACTATTACGGATCATTTGAGCGTCTTTAAAGCTGGCGGTCCCAACCATCTGATCTGG	1130
Qy	367	LeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeuLeuCys	386
		:	
Db	1131	CTTATCTCTCTATTGTTTTCACCTCTCTGCTCAATGCTGTGGCAGAGCTTCTGCAG	1190
Qy	387	PheGlyAspArgGluPheTyrLysAspTrpAsnAlaLysSerValGlyAspTyrTrp	406
		:	
Db	1191	TTTGGACCCGCGAGTTCTACAGAGATGTGTGAATGCTGAGTCTGTCACCTACTTTTGG	1250
Qy	407	ArgMetTrpAsnMetProValHisLysTrpMetValArgHisIleTyrPheProCysLeu	426
		:	
Db	1251	CAGAACTGGAATATCCCGTGCACAAAGTGTGCATCAGACACATCTACAAAGCCTATGCTC	1310
Qy	427	ArgSerLysIleProLysThrLeuAlaIleIleAlaPheLeuValSerAlaValPhe	446
		:	
Db	1311	AGACATGGCAGCAGCAATAATGGGTGGCCAGCAGAGGAGTATTTTTCAGCTCAGCCTTCTTC	1370
Qy	447	HisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIle	466
		:	
Db	1371	CATGAGTACCTAGTACGCGTTCCTCGGATGTTCGCTCTGGGCAATTCACAGCCCATG	1430
Qy	467	MetPheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySerThr	486
		:	
Db	1431	ATGGCTCAGTCCCACTGGCTGGATGTGGGCGGATTC-----TTCCAAGGGAAC	1481
Qy	487	ValGlyAsnMetIlePheThrPheIlePheCysIlePheGlyGlnProMetCysValLeu	506
		:	
Db	1482	TATGGCAATGCAGCTGTGTGG---GTGACACTCATCATTCATGGCAACCGGTGTGCTC	1538
Qy	507	LeuTyrTyrHisAsp	511
		:	
Db	1539	ATGTATGTCACGAC	1553

RESULT 15

```
: Sequence 15, Application US/0157855  
; Publication No. US20020170091A1  
; GENERAL INFORMATION:  
; APPLICANT: Lassarner, Michael W.  
; APPLICANT: Ruezinsky, Diane M.  
; TITLE OF INVENTION: Acyl-CoA:Cholesterol  
Acyltransferase Related Nucleic
```

```

: TITLE OF INVENTION: Acid Sequences
: FILE REFERENCE: 16516.158
: CURRENT APPLICATION NUMBER: US/10/157,855
: CURRENT FILING DATE: 2002-05-31
: PRIOR APPLICATION NUMBER: 09/326,203
: PRIOR FILING DATE: 1999-06-04
: PRIOR APPLICATION NUMBER: 60/088,143
: PRIOR FILING DATE: 1998-06-05
: PRIOR APPLICATION NUMBER: 60/108,389
: PRIOR FILING DATE: 1998-11-12
: NUMBER OF SEQ ID NOS: 46
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 15
: LENGTH: 1766
: TYPE: DNA
: ORGANISM: Rattus sp.
US-10-157-855-15

Alignment Scores:
Pred. No.: 2,7e-76 Length: 1766
Score: 773.50 Matches: 193
Percent Similarity: 53.58% Conservative: 84
Best Local Similarity: 37.33% Mismatches: 171
Query Match: 27.91% Indels: 69
DB: 13 Gaps: 17

US-09-623-514A-2 (1-520) x US-10-157-855-15 (1-1766)

Qy 14 GluAsnGlyGlyGlyGluPheValAspLeuAspArgLeuArgArgLySerArgSer 33
Db 99 CAGGGTGGTAGTGGGCCCATGGTAGACAGAGAGAGGTGCGAGACGCGCTGTGGGCCCC 158
Qy 34 AspSerSerAsnGlyLeuLeuLeuSerGlySerAsnAsnSerProSerAspVal 53
Db 159 GAC-----TTGGGCGCGGGGGTGACGGTCCGGCTCCGGCT---CCGTT 200
Qy 54 GlyAlaProAlaAspValArgAspArgIleAspSerValValAsnAspAlaGlnGly 73
Db 201 CCGGCTCCAGGCCACACCGGGACAAA-----GACCGGCAGACACC 239
Qy 74 ThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGly 93
Db 240 AGCGTG-----GCGGACGCGC 254
Qy 94 GlyGlyGluGlyArgGlyAsnAlaAspAlaThrPheThrTyrArgProSerValProAla 113
Db 255 CACTGGGAGCTGAG-----TCC 272
Qy 114 HisArgArgAlaArgGluSerProLeuSerSerAspAlaIlePheLyGlnSerHisAla 133
Db 273 CAT---CGTCTGCAAGACTCTTTGTCAGCTCAGACAGCGGTTTC---AGCAATTACCGT 326
Qy 134 GlyLeuPheAsnLeuCysValValValIleuIleAlaValAsnSerArgLeuIleIleGlu 153
Db 327 GGTATCTGAATTGGTGGCGGTGATGCTGATCTGATCTAGTAATGCAAGTTATTTTATGAG 386
Qy 154 AsnLeuMetIlyTyrGlyTyrPheLeuIleArgThrAspPheTrpPheSerSerArgSerLeu 173
Db 387 AATCTTATCAATATGGCATCTCTGGTG---CATCCCATCCAGGTGCTCTCTGTTTCTG 443
Qy 174 ArgAsp-----TrpProLeuPheMetCysCysIleSerLeuSerIlePheProLeu 190
Db 444 AAGGACCCCTACAGCTGGCGCTGCCCATGCTTGATCATTTGATGCCAATATCTTTATTGTG 503
Qy 191 AlaAlaPheThrValGluLysLeuValLeuGlnIlyTyrIleSerGluProValIle 210
Db 504 GCTACATTTAGATTGAGAGCGGCTGTCTAGTGGGTGGCCCTGACAGACAGATGGGGCTG 563
Qy 211 PheLeuHisIleIleIleThrMetThrGluValLeuTyrProValTyrValThrLeuArg 230
Db 564 CTGTACATGTGGTTTAACTCGGCCACCAATATCTGCTTCCCGACGAGCTGGCCCTACTG 623
Qy 231 CysAspSerAlaPheLeuSerGlyValThrLeuMetLeuLeuThr---CysIleValTrp 249

```

Db 660 GCAGCTGGCGCTTACTGTTGAGTCTATCACTCCAGTGGGTCCGTTGTTGCTGGCA 719
QY 245 Thr---CysileValtrpLeuLysLeuValSerTyrAlaHisThrSerTyrAspIleArg 263
Db 720 TCATACTCCATCATGTCCTCAAGCTTTATTCCTACCGGAGTCAACCTGGTGGCCGC 779
QY 264 -----SerLeuAlaAsnAlaAlaAsp 270
Db 780 CAGCGAAGGGTCAAGCCAAAGCTGTCTACAGGGAAGAGTCAAGTGGGCTGTCT--- 836
QY 271 LysAlaAsnProGluValSerTyr-----TyrValSerLeuLysSerLeuAlaTyrPhe 288
Db 837 ---GCCACGACCTGTGAGCTATCCAGACACCTGACCTACCGAGATCTCTATTACTTC 893
QY 289 MetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCysIleArgLys 308
Db 894 ATCTTTGCTCCTACTTGTGTTGTTATGAACCTCAACTTTCCTCGTCCCGCCCAATACGAAG 953
QY 309 GlyTrpValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIle 328
Db 954 CGCTTCTCTACGACGAGTCTTTCAGATGCTCTTTTACCGAGCTTCAAGTGGGCGTG 1013
QY 329 IleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeu 348
Db 1014 ATCCACAGTGTGATGCTCTACTATCCACACTCCATGAGCCCTTCAAG---GATARG 1070
QY 349 LeuTyrAla-----IleGluArgValLeuLysLeuSerValProAsnLeuTyrValTrp 366
Db 1071 GACTATTTCACGATCATTCAGCGTCTCTTAAAGCTGGCGTCCCAACCATCTGATCTGG 1130
QY 367 LeuCysMetPheTyrCysPheHisLeuTrpLeuAsnIleLeuAlaGluLeuLysCys 386
Db 1131 CTATCTCTCTTATTTGTTTTCACCTCTCTCTCAATGCTGTGGCAGAGCTTCTGCGAG 1190
QY 387 PheGlyAspArgGluPheTyrLysAspTrpTyrAsnAlaLysSerValGlyAspTyrTrp 406
Db 1191 TTGGAGACCGGAGTCTACAGAGATGTTGGATGCTGAGTCTGCTCACCCTACTTTGG 1250
QY 407 ArgMetTrpAsnMetProValHisLysTrpMetValArgHisIleTyrPheProCysLeu 426
Db 1251 CAGAACTGGAATATCCCGTGCACAAAGTGGTGCATCAGACACTTCTACAAAGCTATGCTC 1310
QY 427 ArgSerLysIleProLysThrLeuAlaIleIleAlaPheLeuValSerAlaValPhe 446
Db 1311 AGACATGGCAGCAGCAATGGGTGGCCAGCAGGAGTATTTTTCACCTCAGCCTTCTTC 1370
QY 447 HisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTyrAlaPheLeuGlyIle 466
Db 1371 CATGAGTACTAGTAGGAGCTTCCCTCGGATGTTCCGCTCTGGGCATTCACAGCCATG 1430
QY 467 MetPheGlnValProLeuValPheIleThrAsnTyrLeuGlnGluArgPheGlySerThr 486
Db 1431 ATGGCTCAGTCCCACTGGCTGGATTGTGGCGGATTC-----TTCCAAGGGAAC 1481
QY 487 ValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCysValLeu 506
Db 1482 TAGGCAATGGACCTGTGTGG---GTGACACTCATCATTTGGCAACCGGTGGTGTGCTC 1538
QY 507 LeuTyrTyrHisAsp 511
Db 1539 ATGATGTCACGAC 1553

RESULT 14

US-10-273-438-9
; Sequence 9, Application US/10273438
; Publication No. US20030072757A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V.
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-Acyltransferase
; FILE REFERENCE: UCAL-105C1P2

; CURRENT APPLICATION NUMBER: US/10/273,438
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US/10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: mus musculus
; US-10-273-438-9
Alignment Scores:
Pred. No.: 2,42e-76 Length: 1650
Score: 773.50 Matches: 191
Percent Similarity: 51.81% Conservative: 81
Best Local Similarity: 36.38% Mismatches: 178
Query Match: 27.91% Indels: 75
DB: 14 Gaps: 15
US-09-623-514A-2 (1-520) x US-10-273-438-9 (1-1650)
QY 8 GlyValThrThrValThrGluAsnGlyGlyGluPheValAspLeuAspArgLeuArg 27
Db 141 GGCTCGCGGTTCCTCCAGGGTGTAGTGGCCCAAGGTAGAGAGGACGAGGTGCGA 200
QY 28 ArgArgLysSerArgSerAspSerSerAsnGlyLeuLeuSerGlySerAspAsn 47
Db 201 GAGCGGCTGTGAGCCCCGAC-----TTGGCCCGGGGGTGACGGCGG 245
QY 48 SerProSerAspValGlyAlaProAlaAspValArgAspArgIleAspSerVal 67
Db 246 GCTCCGCGCTCG-----GCTCCAGCCGATACCCGGGACAA----- 281
QY 68 AsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAspAsnGlyGlyAspAsn 87
Db 281 ----- 281
QY 88 AsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 107
Db 282 ---GACGGGGGACGAGCGTGGCGGCGCTACTGGGATCTGAGGTGC----- 326
QY 108 ArgProSerValProAlaHisArgArgAlaArgGluSerProLeuSerSerAspAlaIle 127
Db 327 -----CATCGCTCTCAAGATTCCTTTGTTCAAGTCTGAGTGGT 365
QY 128 PheLysGlnSerHisAlaGlyLeuPheAsnLeuCysValValLeuLeuAlaValAsn 147
Db 366 TTC---AGCAATATTCGTGTATCTCTGAATTTGGTGTGTGTGTGTGTGTGTGTGT 422
QY 148 SerArgLeuIleIleGluAsnLeuMetLysTyrGlyTrpLeuIleArgThrAspPheTrp 167
Db 423 GCAAGGTATTTTATAGAGAACCTTATCAAGTATGTCATCTCTGGT---GATCCTATCCAG 479
QY 168 PheSerSerArgSerLeuArgAsp-----TrpProLeuPheMetCysIleSer 184
Db 480 GTGGTGTCTCTGTTTGAAGGACCCCTACAGTGGCCCTGCCCCATCGGTGATATTGCA 539
QY 185 LeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeuLysTyrIle 204
Db 540 TCCAAATATTTTGT 599
QY 205 SerGluProValIlePheLeuHisIleIleIleIleIleIleIleIleIleIleIle 224
Db 600 ACAGACAGATGGGCTGCTGTACATGTGGTTAACTGGCCCAATCATTTCTTCCCA 659

```
Db 586 CTGGCGCTGATGGCGGCACACCATCTCTCTCAAGCTCTTCTCTCTACCGC----- 636
Qy 260 TyrAspIleArgSer-----LeuAlaAsnAlaAspLys 271
Db 637 ---GACGTCACACTCATGTGGCGGAGCCAGGCGCAAGGCTGCTCTGCAGGGAAGAG 693
Qy 272 AlaAsnPro-----GluValSerTyr-----TyrValSerLeuLysSer 284
Db 694 GCCAGCAGTGTGTGCCCCGCACACACCGTGAAGTACCGGACAAATCTGACCTACCGCAT 753
Qy 285 LeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAla 304
Db 754 CTCCTACTTCTCTCTCGCCCCACCTGTGCTACGAGCTCAACTTTCCCGCTCTCC 813
Qy 305 CysIleArgLysGlyTyrValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPhe 324
Db 814 CGCATCCGGAAGCGCTTCTCTCGGACGGATCTTGAGATGCTGTCTTCAACCGAGTC 873
Qy 325 MetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeu 344
Db 874 CAGGTGGGCTGATCCAGCAGTGGATGTGCCACCACTCCAGAACTCCATGAAGCCCTTC 933
Qy 345 LysGlyAspLeuLeuTyrAla-----IleGluArgValLeuLysLeuSerValProAsn 362
Db 934 AAG---GACATGGACTACTACGCATCATCGAGCGCTCTCTGAAGTGGCGGTCCCCAAT 990
Qy 363 LeuTyrValTyrLeuCysMetPheTyrCysPheHisLeuTyrLeuAsnIleLeuAla 382
Db 991 CACCTCATCTGGCTCATCTTCTCTACTGGCTCTCTCCACTCTCGCTGAATGCCGTGGCT 1050
Qy 383 GluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTyrTrpAsnAlaLysSerVal 402
Db 1051 GAGCTCATGCACTTGGAGACGGGAGTCTACCGGGACTGGTGAAGTCCGAGTCTGTC 1110
Qy 403 GlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHisIleTyr 422
Db 1111 ACCTACTTCTGGGAGAACTGGAACATCCCTGTGCACAAAGTGGTGCATCAGACACTTCAC 1170
Qy 423 PheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPheLeuVal 442
Db 1171 AAGCCCATCTTCGACGGGGCAGCAGCAAGTGGATGGCCAGCAGGCGTGTCTCGGCC 1230
Qy 443 SerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAla 462
Db 1231 TCGGCTTTCTCCAGCAGTACTGTGTGAGCGCTCTCTCGCAATGTTCCGCTCTGGGCT 1290
Qy 463 PheLeuGlyIleMetPheGlnValProLeuVal---PheIleThrAsnTyrLeuGlnGlu 481
Db 1291 TTCAGGGCATGATGGCTCAGATCCCACTGGCTGTGTTCTGGGCGCGCTTTTTCAGAGGC 1350
Qy 482 ArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGln 501
Db 1351 AACTAT-----GGCAGCCAGCTGTGG---CAGTGCCTCATCATCGGACAG 1395
Qy 502 ProMetCysValLeuLeuTyrTyrHisAsp 511
Db 1396 CCAATAGCGTCTCATGTACGTCCAGCAC 1425
```

RESULT 13

```
US-10-273-438-3
; Sequence 3, Application US/10273438
; Publication No. US20030072757A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V.
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-Acyltransferase
; FILE REFERENCE: UCAL-105GIP2
; CURRENT APPLICATION NUMBER: US/10/273,438
; CURRENT FILING DATE: 2002-10-16
; PRIORITY APPLICATION NUMBER: US/10/040,315
```

```
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: mus musculus
US-10-273-438-3

Alignment Scores:
Pred. No.: 2,42e-76 Length: 1650
Score: 773.50 Matches: 191
Percent Similarity: 51.81% Conservative: 81
Best Local Similarity: 36.38% Mismatches: 178
Query Match: 27.91% Indels: 75
DB: 14 Gaps: 15

US-09-623-514A-2 (1-520) x US-10-273-438-3 (1-1650)
```

```
Qy 8 GlyValThrThrValThrGluAsnGlyGlyGluPheValAspLeuAspArgLeuArg 27
Db 141 GCCTCCGGGTTCCTCCAGGGTGTAGTGGGCCAAGGTAGAGAGACAGAGGTGGCA 200
Qy 28 ArgArgLysSerArgSerAspSerAsnGlyLeuLeuLeuSerGlySerAspAsnAsn 47
Db 201 GACGGCGGTGTGAGCCCGAC-----TTGGGCGCGGGGTGACGCGCG 245
Qy 48 SerProSerAspValGlyAlaProAlaAspValArgAspArgIleAspSerValVal 67
Db 246 GCTCGCGGTCCG-----GCTCCAGCCCATACCGGACAAA----- 281
Qy 68 AsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAspAsnAsnGlyGlyGlyAspAsn 87
Db 281 ----- 281
Qy 88 AsnGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 107
Db 282 ---GACGGCGGACGACGCGTGGCGCGCTACTGGGATCTGAGGTGC----- 326
Qy 108 ArgProSerValProAlaHisArgAlaArgAlaArgGluSerProLeuSerSerAspAlaIle 127
Db 327 -----CATGCTCTGCAAGATTCTTTGCTCAGCTCAGACAGTGGT 365
Qy 128 PheLysGlnSerHisAlaGlyLeuPheAsnLeuCysValValValIleAlaValAsn 147
Db 366 TTC---ACCAATTATGCTGGTATCTGAAATGGTGGTGGTGGTGGTGGTGGTGGTGGT 422
Qy 148 SerArgLeuIleIleGluAsnLeuMetLysTyrGlyTyrLeuIleArgThrAspPheTrp 167
Db 423 GCAAGGTTATTTTAGAGAACCTTATCAAGTATGCAATCCCTCTGGT---GATCCTATCCAG 479
Qy 168 PheSerSerArgSerLeuArgAsp-----TrpProLeuPheMetCysCysIleSer 184
Db 480 GTGGTGTCTCTGTTTGTGAAGGACCCCTACAGCTGGCTGGCTGGCTGGCTGGCTGGCT 539
Qy 185 LeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeuGlnLysTrpIle 204
Db 540 TCCAATATTTTGTGGCTGCAATTCAGATTGAGAAGCGCTGGAGCGCTGGAGTGGTGG 599
Qy 205 SerGluProValIlePheLeuHisIleIleIleIleIleIleIleIleIleIleIleIle 224
Db 600 ACAGACGAGATGGGCTGCTGCTACATGTTACCTGGCCACACATATTGCTTCCCA 659
Qy 225 ValTyrValThrLeuArgCysSerAlaPheLeuSerGlyValThrLeuMetLeuLeu 244
```



```
Db 72 GTCGATATGGAATATGTCGATCCGAGATATGAAGATCAACCTTTAAAGCTA 131
Qy 286 AlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyrProArgSerAlaCys 305
Db 132 GTGTACTTCATGTTGGCCCAACACTTGTACCAGCAACTTATCTCAACTACATGT 191
Qy 306 IleArgLysGlyTyrPValAlaArgGlnPheAlaLysLeuValIlePheThrGlyPheMet 325
Db 192 ATTAAGAAGGGTGGTGACCACTCAATAAAGTCGCTGGNNNNNACAGGCTTGATG 251
Qy 326 GlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSerLysHisProLeuLys 345
Db 252 GGCTTCATATATGAGCAATATATAACCAATTTGTAAGATTCCTCAACATCCACTGAAA 311
Qy 346 GlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerValProAsnLeuTyrVal 365
Db 312 GGGANNNNNGAATGCTATAGAAGAGTCTTAACATCTCAGTGCCCAACATATATGTA 371
Qy 366 TrpLeuCysMetPheTyrCysPhePheHisLeuTrpLeuAsnIleLeuAlaGluLeu 385
Db 372 TGGCTTTGCATGTTCTATGTCNNNNNCATTTATGGCTGAACATTTAGCTGAACCTCCTC 431
Qy 386 CysPheGlyAspArgGluPheTyrLysAspTrpTrpAsnAlaLysSerValGlyAspTyr 405
Db 432 TGTTTCGCTGACCGTGAATCTATAGGACTGGTGAATGCCAAACTGTTGAAGAGTAC 491
Qy 406 TrpArgMetTrpAsnMetProValHisLysTrpMetValArgHisIleTyrPheProCys 425
Db 492 TGGAGGATGTGAACATGCGCTGTCATAGTGATGATCATCAGACACATATATTTCCATGT 551
Qy 426 LeuArgSerLysIleProLysThrLeuAlaIleIleLeuAlaPheLeuValSerAlaVal 445
Db 552 ATAAGGAAAGGCTTTCCAGGGGTAGTATTAATCTCGTTCTGTTTTCAGCTGTA 611
Qy 446 PheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeuTrpAlaPheLeuGly 465
Db 612 TTCATGAGATATATATGCGGTGCCGTCACATTTTCAAAATTTCTGGCATTTTCTGGG 671
Qy 466 IleMetPheGlnValProLeuValPheIlePheTrpAsnTyrLeuGlnArgPheGlySer 485
Db 672 ATCATGTTTCAGATACCGTGGTATTTCTTGACAGATATCTCCATGTCAGTTCACGAT 731
Qy 486 Thr---ValGlyAsnMetIlePheTrpPheIlePheCysIlePheGlyGlnProMetCys 504
Db 732 GTAATGTTGGCAACATGATATTTTGGTTC---TTCAGTATAGTCGACAGCCCGATGTT 788
Qy 505 ValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520
Db 789 GTCCTTCTATACTACCATGACGTGATGAACAGCGAGGCCCGCCAGCAAGT 836
```

RESULT 11

US-10-278-733-2

; Sequence 2, Application US/10278733

; Publication No. US20030100480A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Steven

; APPLICANT: Chen, Hubert

; TITLE OF INVENTION: Methods and compositions for modulating

; TITLE OF INVENTION: sebaceous glands

; FILE REFERENCE: UCAL-105CIP4

; CURRENT APPLICATION NUMBER: US/10/278,733

; PRIOR FILING DATE: 2002-10-21

; PRIOR APPLICATION NUMBER: 10/040,315

; PRIOR FILING DATE: 2001-10-29

; PRIOR APPLICATION NUMBER: 09/339,472

; PRIOR FILING DATE: 1999-06-23

; PRIOR APPLICATION NUMBER: 60/107,771

; PRIOR FILING DATE: 1998-11-09

; PRIOR APPLICATION NUMBER: PCT/US98/17883

; PRIOR FILING DATE: 1998-08-28

; PRIOR APPLICATION NUMBER: 09/103,754

```
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1467)
; OTHER INFORMATION: Homo sapiens diacylglycerol O-acyltransferase
; OTHER INFORMATION: homolog 1
US-10-278-733-2
```

Alignment Scores:

Pred. No.:	1,9e-78	Length:	1467
Score:	791.50	Matches:	194
Percent Similarity:	53.14%	Conservative:	77
Best Local Similarity:	38.04%	Mismatches:	174
Query Match:	28.56%	Indels:	65
DB:	14	Gaps:	17

US-09-623-514A-2 (1-520) x US-10-278-733-2 (1-1467)

```
Qy 25 ArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeuSerGlySer 44
Db 22 CGGCGCGGAGACAGGTCGCGCCCTCGAGCCACCGCGGC----- 63
Qy 45 AspAsnAsnSerProSerAspAspValGlyAlaProAlaAspValArgAspArgIleAsp 64
Db 64 -----GGCGGGCTCGCGCGCGGAGAGGAG----- 90
Qy 65 SerValValAsnAspAspAlaGlnGly---ThrAlaAsnLeuAlaGlyAspAsnGly 83
Db 91 -----GTGCGGAGCGCGCTGCGGCGCCCGCACCTGGAGCGCGGGGAGCCGCCAGCC 144
Qy 84 GlyGlyAspAsnAsnGlyGlyArgGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 103
Db 145 CCGGCCCCCAACAGGAGCGGAGCGCGCGCTGGGCGGCGGCGGCGGCGGAGCTGAGTGC 204
Qy 104 ThrPheThrTyrArgProSerValProAlaHisArgAlaArgGluSerProLeuSer 123
Db 205 -----CATGCGCTGCAGGATCTTTATTTCAGC 231
Qy 124 SerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysValValValLeu 143
Db 232 TCTGACAGTGGCTTC---AGCAACTACCGTGGCATCTGAACTGGTGTGTGTGTGTGTGTGT 288
Qy 144 IleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTyrTrpLeuIleArg 163
Db 289 ATCTTGAGCAATGCGCGTTATTTCTGGAGAACCTCATCAAGTATGTCATCTCTGCTGGT 345
Qy 164 ThrAspPheTrpPheSerSerArgSerLeuArgAsp-----TrpProLeuPheMet 180
Db 346 GACCCCATCCAGGTGGTTTCTCTTCTTCTGAAGATCCCTATAGTGGCGCGCCCATGCG 405
Qy 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200
Db 406 CTGCTTATTTGGGCGCAATGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 465
Qy 201 GlnLysTyrIleSerGluProValIlePheLeuHisIleIleIleIleIleIleIleIle 220
Db 466 GTGGTGGCTGCGGAGGAGCGGCGGAGCTGCTGTGCGACGTCGCGCAACCTGGCCACCAT 525
Qy 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
Db 526 CTGTGTTTCCAGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 585
Qy 241 LeuMetLeuLeuThr---CysIleValTrpLeuLysLeuValSerTyrAlaHisThrSer 259
Db 586 CTGGCGGTGTGGCGCACACCATCTCTCTCTCAAGCTCTCTCTCTCTCTCTCTCTCTCT 636
Qy 260 TyrAspIleArgSer-----LeuAlaAsnAlaAlaAspLys 271
```

[illegible]

; APPLICANT: Farese, Robert V
 ; APPLICANT: Cases, Sylvaine
 ; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and
 ; FILE REFERENCE: UCAL-105CIP3
 ; CURRENT APPLICATION NUMBER: US/10/223,076
 ; PRIOR FILING DATE: 2001-10-29
 ; PRIOR APPLICATION NUMBER: 10/040,315
 ; PRIOR FILING DATE: 2001-10-29
 ; PRIOR APPLICATION NUMBER: 09/339,472
 ; PRIOR FILING DATE: 1999-06-23
 ; PRIOR APPLICATION NUMBER: 60/107,771
 ; PRIOR FILING DATE: 1998-11-09
 ; PRIOR APPLICATION NUMBER: PCT/US98/17883
 ; PRIOR FILING DATE: 1998-08-28
 ; PRIOR APPLICATION NUMBER: 09/103,754
 ; PRIOR FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 1446
 ; TYPE: DNA
 ; ORGANISM: Brassica napus
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (82)...(1107)
 US-10-223-076-6

Alignment Scores:

Pred. No.: 2,2e-195 Length: 1446
 Score: 1833.00 Matches: 342
 Percent Similarity: 95.6% Conservative: 10
 Best Local Similarity: 92.93% Mismatches: 16
 Query Match: 66.15% Indels: 1
 DB: 14 Gaps: 0

US-09-623-514A-2 (1-520) x US-10-223-076-6 (1-1446)

QY 153 GluAsnLeuMetLysTyrglyTrpLeuIleAthrAspPheTrpPheSerSerArgSer 172
 DB 2 GAAAATCTCATGAAGTACGGTGGTGTGATCAGACTGATTTCTGGTTTGTAGTTCAACGTCG 61
 QY 173 LeuArgAspTrpProLeuPheMetCysIleSerLeuSerIlePheProLeuAlaIa 192
 DB 62 CTGCGAGATG-CGCGTTTTCATGTGTCTCTCCCTTCAATCTTCTTGGTGGTCC 120
 QY 193 PheThrValGluLysLeuValLeuGlnLysTyriIleSerGluProValIlePheLeu 212
 DB 121 TTTACCGTCGGAATATAGTACTCAGAAATGATATCTGAACCTGTGTCTATCTTTCTT 180
 QY 213 HisIleIleIleThrMetThrGluValLeuTyriProValTyriValThrLeuArgCysAsp 232
 DB 181 CATGTTATATACCATCACCAGGCTCTGTATCCAGTCTATGTCACCTTAAGGTGTGAT 240
 QY 233 SerAlaPheLeuSerGlyValThrLeuMetLeuThrCysIleValTrpLeuLysLeu 252
 DB 241 TCTGCTCTTATCAGGTGACAGCTGTGATGCTCTCACTTGCATTTGTGGCTGAAGTTG 300
 QY 253 ValSerTyriAlaHisThrSerTyriAspIleArgSerLeuAlaAsnAlaAlaAspLysAla 272
 DB 301 GTTCTTACGCTCATCTAATATAGACATAGACAGCTAGCTAGTAAATCATCTGATAAGGCC 360
 QY 273 AsnProGluValSerTyriTyriValSerLeuLysSerLeuAlaTyriPheMetValAlaPro 292
 DB 361 AATCCTGAAGTCTCTACTATGTAGCTTGAAGAGCTTGGCTTATTCATCTGCTGCTCC 420
 QY 293 ThrLeuCysTyriGlnProSerTyriProArgSerAlaCysIleArgLysGlyTrpValAla 312
 DB 421 ACATTGTGTATCATGCCAAGCTATCCAGCTTCTCCATGTATCCGGAAGGTTGGTGGCT 480
 QY 313 ArgGlnPheAlaLysLeuValIlePheThrGlyPheMetGlyPheIleGluGlnTyri 332
 DB 481 GGTCAATTTGCAAACTGGTCTATATTTCACTGGACTCATGGGATTTATAATAGACAATAT 540

QY 333 IleAsnProIleValArgAsnSerLysHisProLeuLysGlyAspLeuLeuTyriAlaIle 352
 DB 541 ATAAATCTATTGTTAGGAACCTCAAAGCATCTCTGAAAGGGACCTTCTATATGCTATT 600
 QY 353 GluArgValLeuLysLeuSerValProAsnLeuTyriValTrpLeuCysMetPheTyriCys 372
 DB 601 GAAAGAGTGTGAAGCTTTTCAGTTTCCAAATCTATATGTGTGGCTCTGCTATGCTTCTACTG 660
 QY 373 PhePheHisLeuTyriLeuAsnIleLeuAlaGluLeuLeuCysPheGlyAspArgGluPhe 392
 DB 661 TTCCTCCACCTTGTGTTAAACATATTTGGCAGAGCTCTCTGCTTGGGGGACCGTGAATTC 720
 QY 393 TyrLysAspTrpTrpAsnAlaLysSerValGlyAspTyriTrpArgMetTrpAsnMetPro 412
 DB 721 TACAAAGATTGGTGAATGCAAAAGCGTTGGAGATTTATGGAGAAATGCGAATATGCGCT 780
 QY 413 ValHisLysTrpMetValArgHisIleTyriPheProCysLeuArgSerLysIleProLys 432
 DB 781 GTTCACAAATGGATGGTTGCGACATGTATCTTCTAGTCTCTGAGTCTTTCATGAGTTATGCA 840
 QY 433 ThrLeuAlaIleIleIleAlaPheLeuValSerAlaValAlaPheHisGluLeuCysIleAla 452
 DB 841 GTACCCGCCATTATCTTCTTCTAGTCTCTGAGTCTTTCATGAGTTATGCAATGCA 900
 QY 453 ValProCysArgLeuPheLysLeuTrpAlaPheLeuGlyIleMetPheGlnValProLeu 472
 DB 901 GTTCTTGGCGCTCTCTCAATCTATGGGCTTTCATGGGAATATGTTTCAGTCCCTTTG 960
 QY 473 ValPheIleThrAsnTyriLeuGlnGluArgPheGlySerThrValGlyAsnMetIlePhe 492
 DB 961 GTCTTTATCACAAATCTTTTACAAAGAAAGTTTGGCTCCATGGTGGGAAACATGATCTTT 1020
 QY 493 TrpPheIlePheCysIlePheGlyGlnProMetCysValLeuLeuTyriTyriHisAspLeu 512
 DB 1021 GGTTCAGCTTCTTGCATTTTCGGACACCGATGTGGGCTTCTTTATTACCATGACCTG 1080
 QY 513 MetAsnArgLysGlySerMetSer 520
 DB 1081 ATGAACGCAAAAGGATCCATGCTCC 1104

RESULT 9

; Sequence 17, Application US/10223076
 ; Publication No. US20030074695A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Farese, Robert V
 ; APPLICANT: Cases, Sylvaine
 ; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and
 ; FILE REFERENCE: UCAL-105CIP3
 ; CURRENT APPLICATION NUMBER: US/10/223,076
 ; CURRENT FILING DATE: 2001-10-29
 ; PRIOR APPLICATION NUMBER: 10/040,315
 ; PRIOR FILING DATE: 2001-10-29
 ; PRIOR APPLICATION NUMBER: 09/339,472
 ; PRIOR FILING DATE: 1999-06-23
 ; PRIOR APPLICATION NUMBER: 60/107,771
 ; PRIOR FILING DATE: 1998-11-09
 ; PRIOR APPLICATION NUMBER: PCT/US98/17883
 ; PRIOR FILING DATE: 1998-08-28
 ; PRIOR APPLICATION NUMBER: 09/103,754
 ; PRIOR FILING DATE: 1998-06-24
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 17
 ; LENGTH: 1572
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 US-10-223-076-17

Alignment Scores:

Pred. No.: 3.96e-133 Length: 1572

; COLLECTION NO: 0320
; GENERAL INFORMATION:


```

; SEQ ID NO 10
; LENGTH: 2090
; TYPE: DNA
; ORGANISM: Tropaeolum majus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (171)...(1727)
US-10-223-076-10

Alignment Scores:
Pred. No.: 2,13e-200 Length: 2090
Score: 1880.00 Matches: 370
Percent Similarity: 78.11% Conservative: 44
Best Local Similarity: 69.81% Mismatches: 88
Query Match: 67.85% Indels: 28
DB: 14 Gaps: 10

US-09-623-514A-2 (1-520) x US-10-223-076-10 (1-2090)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGlyGluPhe
Db 171 ATGGCGGTGGCAGAGTCGTCACAGAACCGACACA---ACCATGAGTGTGTACGCGCAGAC--- 224

QY 21 ValAspLeuAspArgLeuArgArgGlySerArgSerAsp
Db 225 TCGGATCTCAACAATTCCGTGAGAAGGAACCGAGTTCCTCCCGTGATGAACCTTCGTGCG 284

QY 37 AsnGlyLeuLeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaPro 56
Db 285 TCCGCTTTTACA-----TCCACCANTGGCGTACCGCGG-----ACTGCCACGCTG 329

QY 57 AlaAspValArgAspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsn 76
Db 330 GCTGAGAATCGTGACCAAGATCGGTAGGGGCTATGGAGAACGACAGGATCGGTCAAC 389

QY 77 LeuAlaGlyAspAsnAsnGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlu 96
Db 390 TTAATTGGA-----AATGGTGGAGCGGTGGTTATCGGGAATGA 428

QY 97 GlyArg-----GlyAsnAlaAspAlaThrPheThrTyrArgProSerValProAlaHis 114
Db 429 GAGAAACAGGTAGGGGAGACTGATACGATTCACTTACCGGCTTCGTTTCGGCTCAT 488

QY 115 ArgArgAlaArgGluSerProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGly 134
Db 489 CGGAGGGTGGAGGAGAGTCTCTTGTAGTCTGATGCAATCTTAAACAGAGCCATCGGGT 548

QY 135 LeuPheAsnLeuCysValValValIleAlaValAsnSerArgLeuIleIleGluAsn 154
Db 549 TTTATTCAACTTGTGTATAGTAGTCTCATTCAGTAGTAACAGTAGTGGCTTATCATCGAAA 608

QY 155 LeuMetLysTyrGlyTrpLeuIleArgThrAspPheTrpPheSerSerArgSerLeuArg 174
Db 609 CTTATGAGTAGTGGTGGTTCATCGATCACTCGGTTTCGGTTTAACTCAAGATCATCGGGT 668

QY 175 AspTrpProLeuPheMetCysCysIleSerLeuSerIlePheProLeuAlaIaPheThr 194
Db 669 GATTGGTCCATCTTTATGTGCTGTCTTACACTCCCAATTTTCCCACTTGTGCTGTTTTATT 728

QY 195 ValGluLysLeuValLeuGlnLysTyrIleSerGluProValIlePheLeuHisIle 214
Db 729 GTTGAAGAGCTGGTGCAGCGAAATCATATCTGAACCTTGTCTGTTCTCTTCATGTA 788

QY 215 IleIleThrMetThrGluValLeuTyrProValTyrValThrLeuArgCysAspSerAla 234
Db 789 ATCGGTTCTACCGCTGCAGCTTTTATATCCAGTATTGTGATCTTACGCTGTGATCGGG 848

QY 235 PheLeuSerGlyValThrLeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSer 254
Db 849 TATATGCTGGTGGTATTGATGCTCTTTGGTGTGATATGTTGGTGGTGGTGGTGGTGA 908

QY 255 TyrAlaHisThrSerTyrAspIleArgSerLeuAlaAsnAlaAlaAspLys-----Ala 272

```

```

: NAME/KEY: CDS
: LOCATION: (1)...(1512)
US-10-223-076-8

Alignment Scores:
Pred. No.: 7,55e-249 Length: 1512
Score: 2309,50 Matches: 442
Percent Similarity: 88.10% Conservative: 17
Best Local Similarity: 84.84% Mismatches: 43
Query Match: 83.35% Indels: 19
DB: 14 Gaps: 5

US-09-623-514R-2 (1-520) x US-10-223-076-8 (1-1512)

Qy 1 MetAlaIleLeuAspSerAlaGlyValThrThr---ValThrGluAsnGlyGlyGlu 19
Db 1 ATGGCGATTTGGATTTCTGGAGGCGTCGCTACCGCGGAGGAGACGGC----- 51

Qy 20 PheValAspLeuAspArgLeuArgArgArgLysSerArgSerAspSerAsnGlyLeu 39
Db 52 GTCGCGGATCTCGACAGGCTCCACCGTCGTAATTCGAGTTCCGATTTCTTCCACGGACTC 111

Qy 40 LeuLeuSerGlySerAspAsnSerProSerAspValGlyAlaProAlaAspVal 59
Db 112 CTC-----TCCGATACTTCCCGTCGCGACGATGTGGAGTCGCGCGCGCGAA 159

Qy 60 ArgAspArgIleAspSerValAlaAsnAspAlaGlnGlyThrAlaAsnLeuAlaGly 79
Db 160 AGGATCGGGTTGATTCGCTGCGGAGGAGGCTCAGGGAACAGCGAATTTAGCT--- 216

Qy 80 AspAsnAsnGlyGlyGlyAspAsnAsnGlyGlyArgGlyGlyGlyGlyGly 99
Db 217 -----GCGGAGATGCGGAAACTAGGGAATCCGCGGAGGC----- 252

Qy 100 AsnAlaAspAlaThrPheThrArgProSerValProAlaHisArgAlaArgGlu 119
Db 253 -----GATGAAGTTTAGTATGATCGACCGTCGCTCCAGCTATCCGAGGACGAGGAG 306

Qy 120 SerProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCys 139
Db 307 AGTCCTCTCAGCTCCGACGCTATCTTCAACAAGACCATGAGGATTTGTTCAACCTCTGT 366

Qy 140 ValValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGly 159
Db 367 GTAGTTGTTCTGTGTGTAAACAGTAGACTCATCATCGAAAAACCTCATGAAGTATGGT 426

Qy 160 TrpLeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTTPProLeuPhe 179
Db 427 TGGTTGATCAGACTGATTTTGGTTAGTTTACATCCCTTACGAGACTGCGCGCTTTTC 486

Qy 180 MetCysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuVal 199
Db 487 ATGGTGTCTCTTTTCCATTTCCGTTTCCGCTGCTTCCGCTGCTGCGGTCGAGAAATGTA 546

Qy 200 LeuGlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleThrMetThr 219
Db 547 CTTTCAGAAATTCATATCATGACCTGTGTCATCATCTTCTGATGTCATTAACCATGACA 606

Qy 220 GluValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyVal 239
Db 607 GAGTCTGTATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 666

Qy 240 ThrLeuMetLeuLeuThrCysIleValTrpLeuLysLeuValSerTyrAlaHisThrSer 259
Db 667 ACGTTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 726

Qy 260 TyrAspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrTyr 279
Db 727 TACGACATAAGAACCCCTGGCAATTCAGCTGATAAGGTCGATCCTGAAATCTCTCTACTAT 786

Qy 280 ValSerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSer 299
Db 787 GTTAGCTTGAGAGCTTGGCGTATTTTCATGGTTGCTCCACACATGTTTATCAGCCAGGC 846

Qy 300 TyrProArgSerAlaCysIleArgLysGlyTyrPValAlaArgGlnPheAlaLysLeuVal 319
Db 847 TATCCACGTTCTCCATGTATCCGGAAGGTTGGTGGCTGCTCAACTTGCAAACTGGTC 906

Qy 320 IlePheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsn 339
Db 907 ATATTCTACTGGACTCATGGGATTTAATAGAGCAATATATAATCTCTATTTGTTAGAAC 966

Qy 340 SerLysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSer 359
Db 967 TCAAGCATCTCTGAAAGGGGACCTTCTATATGCTATTGAAAGAGTGTGAAGCTTTCA 1026

Qy 360 ValProAsnLeuTyrValTrpLeuCysMetPheTyrCysPheHisLeuTrpLeuAsn 379
Db 1027 GTTCCAAATCTATATGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1086

Qy 380 IleLeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTTPTrpAsnAla 399
Db 1087 ATATTGCGAGAGCTCTCTGCTTCCGGGACCGTGAATTTCTACAAAGATTTGTTGGAATGCA 1146

Qy 400 LysSerValGlyAspTyrTTPArgMetTrpAsnMetProValHisLysTTPMetValArg 419
Db 1147 AARAGCCTTGGAGATTTTGGAGATTTGGAATATGCTGTTCAAAATGGATGGTTCGA 1206

Qy 420 HisIleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAla 439
Db 1207 CATGTATACTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1266

Qy 440 PheLeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLys 459
Db 1267 TTCTAGTCTCTGCGAGCTTTTCATGAGTATGATGATGATGATGATGATGATGATGATGATGAT 1326

Qy 460 LeuTrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeu 479
Db 1327 CTATGGCTTTTCATGGGAATTTGTTTCAGGTCCTTTGGTCTTTATACAAACTTTTA 1386

Qy 480 GlnGluArgPheGlySerThrValGlyAsnMetIlePheTrpPheIlePheCysIlePhe 499
Db 1387 CAAGAAAGGTTTGGCTCCATGCTGGGAAACATGATCTTTGTTGCTGCTGCTGCTGCTGCTGCT 1446

Qy 500 GlyGlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMet 519
Db 1447 GGACAAACCGATGTGTGGGCTTCTTTATACCATGATGATGATGATGATGATGATGATGATGATG 1506

Qy 520 Ser 520
Db 1507 TCC 1509

RESULT 5
US-10-223-076-10
: Sequence 10, Application US/10223076
: Publication No. US20030074695A1
: GENERAL INFORMATION:
: APPLICANT: Farese, Robert V
: APPLICANT: Cases, Sylvaine
: TITLE OF INVENTION: Plant Diacylglycerol O-transferase and
: TITLE OF INVENTION: Uses thereof
: FILE REFERENCE: UCAL-105CIP3
: CURRENT APPLICATION NUMBER: US/10/223,076
: PRIOR FILING DATE: 2001-10-29
: PRIOR APPLICATION NUMBER: 10/040,315
: PRIOR FILING DATE: 2001-10-29
: PRIOR APPLICATION NUMBER: 09/339,472
: PRIOR FILING DATE: 1999-06-23
: PRIOR APPLICATION NUMBER: 60/107,771
: PRIOR FILING DATE: 1998-11-09
: PRIOR APPLICATION NUMBER: PCT/US98/17883
: PRIOR FILING DATE: 1998-08-28
: PRIOR APPLICATION NUMBER: 09/103,754
: PRIOR FILING DATE: 1998-06-24
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: FastSEQ for Windows Version 4.0
```


US-10-157-855-1

Alignment Scores:

Pred. No.: 1.76e-300 Length: 1942
Score: 2771.00 Matches: 520
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 13 Gaps: 0

US-09-623-514A-2 (1-520) x US-10-157-855-1 (1-1942)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrValThrGluAsnGlyGlyGluPhe 20
DB 237 ATGGCGATTTTGGATTCTCGCTGGCGTTACTACGGTGACGGAGAACGGTGGCGGAGATTC 296
QY 21 ValAspLeuAspArgLeuArgArgLysSerArgSerAspSerSerAsnGlyLeuLeu 40
DB 297 GTCCGATCTGTATAGCTTCGTCGACGGAATCGAGATCGGATTTCTTCTAACGGACTTCTT 356
QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyAlaProAlaAspValArg 60
DB 357 CTCCTCTGGTTCCGATAATAATCTCTTCGGATGATGTTGGAGCTCCGCCGACGTTAGG 416
QY 61 AspArgIleAspSerValValAsnAspAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
DB 417 GATCGGATTTGATTCGGTTGTTAACGATGACGCTCAGGGAACAGCCAAATTTGGCCGGAGAT 476
QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGly 100
DB 477 AATAACGCTGGTGGCGATTAATACGGTGGTGGAGAGCGCGGAGGAAGGAAGGAAGGA 536
QY 101 AlaAspAlaThrPheThrTyArgProSerValProAlaHisArgAlaArgGluSer 120
DB 537 GCCGATGTACGTTTACGTATACGCGCGTCCGCTCCAGCTCATCGAGGCGGAGAGAGT 596
QY 121 ProLeuSerSerAspAlaIlePheGlyGlnSerHisAlaGlyLeuPheAsnLeuGlyVal 140
DB 597 CCACCTAGCTCCGACGCAATCTTCAACAGAGCCATCGCGGATTTATCAACCTCTCGTGA 656
QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyGlyTyr 160
DB 657 GTAGTTCTATTGCTGTAAACAGTAGACTCATCATCGAATAATCTTATGAAGTATGGTTGG 716
QY 161 LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTyrProLeuPheMet 180
DB 717 TTGATCAGAACGGATTTCTGTTTGTAGTTCAAGATCGCTGCGAGATTTGGCCGCTTTTCATG 776
QY 181 CysCysIleSerLeuSerIlePheProLeuAlaIlePheThrValGluLysLeuValLeu 200
DB 777 TGTGTATATCCCTTTCGATCTTCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCTTTCGCT 836
QY 201 GlnLysTyIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220
DB 837 CAGAAATACATATCAGAACCTGTTGTCATCTTCTTCATATATATATACCCATGACAGAG 896
QY 221 ValLeuTyProValTyValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
DB 897 GTTTTGTATCCAGTTTACGTCACCTTAAGGTGTCATCTGCTTTTATCAGGTGTCAT 956
QY 241 LeuMetLeuLeuThrCysIleValTyIlePheLeuLysLeuValSerTyAlaHisThrSerTy 260
DB 957 TTGATGCTCCCTACCTTTCGATTTGTTGGTAAAGTTGGTTTCTTATGCTCATAGCTAT 1016
QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyVal 280
DB 1017 GACATAGATCCCTAGCCATATGACGTATAGGCCAATCTGAGTCTCTACTACGTT 1076
QY 281 SerLeuLysSerLeuAlaTyPheMetValAlaProThrLeuCysTyArgGlnProSerTy 300
DB 1077 AGCTTGAAGAGCTTGGCATATTTTCATGGTCGCTCCACATGTTGTTATCAGCCAGTTAT 1136
QY 301 ProArgSerAlaCysIleArgLysGlyTyrPValAlaArgGlnPheAlaLysLeuValIle 320

DB 1137 CCACGTTCTGCTATATACGGAAGGTTGGTGGCTCGTCAATTTGCAAACTGGTCATA 1196
QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyIleAsnProIleValArgAsnSer 340
DB 1197 TTCACCGGATTCATGGGATTTATTAAGAACAAATATATAATCTTATGTCAGGAACATCA 1256
QY 341 LysHisProLeuLysGlyAspLeuLeuTyAlaIleGluArgValLeuLysLeuSerVal 360
DB 1257 AAGCATCCCTTTGAAAGGCGATCTCTATATGCTATTGAAAGAGTGTTCGAAGCTTTCAGTT 1316
QY 361 ProAsnLeuTyValTrpLeuCysMetPheTyCysPhePheHisLeuTrpLeuAsnIle 380
DB 1317 CCAAAATTAATATGTGTGGCTCTGCTACTGTCTTCTTCCACCTTTGGTTTAAACATA 1376
QY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyLysAspTyrTrpAsnAlaLys 400
DB 1377 TTGGCAGAGCTTCTGCTTCGGGATCGTGAATCTTACAAGATTTGGTGGAAATGCAANA 1436
QY 401 SerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
DB 1437 AGTGTGGGAGATTACTGGAGAATGTGGAATATGCTGTTTCAATAATGGATGGTTCGACAT 1496
QY 421 IleTyPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleIleAlaPhe 440
DB 1497 ATATACCTCCGCTGCTTCGCGAGCAAGATACCAAGACACTCGCCATTTATCATTTGCTTTC 1556
QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460
DB 1557 CTAGTCTCTGAGCTTCTTCATGAGCTATGATCGAGCTTCTTCTGCTCTCTTCAAGCTA 1616
QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyLeuGln 480
DB 1617 TGGGCTTTTCTTGGGATTTATCTTTCAGGTGCTTGGTCTTTCATCACAACATCTACAG 1676
QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheThrPheIlePheCysIlePheGly 500
DB 1677 GAAAGGTTTGGCTCAACGGTGGGAAACATGATCTTCTGGTTCATCTTCTGCTTTCGGA 1736
QY 501 GlnProMetCysValLeuLeuTyTyHisAspLeuMetAsnArgLysGlySerMetSer 520
DB 1737 CAACCGATGCTGTGCTTCTTTATACCACGACCTGATGAACCCGAAAGGATCGATGTCA 1796

RESULT 3
US-10-223-076-4
; Sequence 4, Application US/10223076
; Publication No. US20030074695A1
; GENERAL INFORMATION:
; APPLICANT: Faresse, Robert V
; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and
; FILE OF INVENTION: Uses Thereof
; FILE REFERENCE: UCAL-105CIP3
; CURRENT APPLICATION NUMBER: US/10/223,076
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1537
; TYPE: DNA
; ORGANISM: Brassica napus
; FEATURE:
; NAME/KEY: CDS

```

; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (139)...(1701)
; US-10-223-076-2

Alignment Scores:
Pred. No.: 1.7e-300 Length: 1904
Score: 2771.00 Matches: 520
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 14 Gaps: 0

US-09-623-514A-2 (1-520) x US-10-223-076-2 (1-1904)

QY 1 MetAlaIleLeuAspSerAlaGlyValThrThrValThrGluAsnGlyGlyGluPhe 20
DB 139 ATGGCGATTTCGATCTGCTGCGCGTTACTACGGTGACGGAGAACGGTGGCGGAGAGTTC 198

QY 21 ValAspLeuAspArgLeuArgArgGlySerArgSerArgSerSerAsnGlyLeuLeu 40
DB 199 GTTCGATCTGTAGAGCTTCGTCGACGGAATCGAGATCGGATCTCTTCAACGCGACTCTT 258

QY 41 LeuSerGlySerAspAsnAsnSerProSerAspValGlyValAlaProAlaAspValArg 60
DB 259 CTCCTGTTGTCGGATAAATCTCTCTCGGATGATGTTGGAGCTCCCGCGACGTTAGG 318

QY 61 AspArgIleAspSerValValAsnAspAlaGlnGlyThrAlaAsnLeuAlaGlyAsp 80
DB 319 GATCGGATTGATTCGTTTAAACGATGACGCTCAGGGAACAGCCAAATTTGGCCGGAGAT 378

QY 81 AsnAsnGlyGlyGlyAspAsnAsnGlyGlyGlyArgGlyGlyGlyGlyGlyGlyGlyAsn 100
DB 379 AATAACGGTGGTGGCGATAAATACGGTGGTGGACAGGCGCGGAGAGAGAGAGAAC 438

QY 101 AlaAspAlaThrPheThrTyrArgProSerValProAlaHisArgAlaArgGluSer 120
DB 439 GCGGATGCTACGTTTACGCTATCGACCGTGGTTCGAGCTTCATCGGAGCGCGAGAGAGT 498

QY 121 ProLeuSerSerAspAlaIlePheLysGlnSerHisAlaGlyLeuPheAsnLeuCysVal 140
DB 499 CCACCTTAGCTCCGACGCAATCTTCAACAGACGACATGCGGATATTCACCTCTGTGTA 558

QY 141 ValValLeuIleAlaValAsnSerArgLeuIleIleGluAsnLeuMetLysTyrGlyTyr 160
DB 559 GTAGTCTTATTGCTGTAACAGTAGACTCATCATCGAAATCTTATGAAGTATGGTTGG 618

QY 161 LeuIleArgThrAspPheThrPheSerSerArgSerLeuArgAspTrpProLeuPheMet 180
DB 619 TTGATCAGAACGGATTTCTGGTTTATGTTCAAGATCGCTCGAGATTTGGCGCTTTTCATG 678

QY 181 CysCysIleSerLeuSerIlePheProLeuAlaAlaPheThrValGluLysLeuValLeu 200
DB 679 TGTGTTATATCCCTTTCGATCTTCTTGGCTGCTTTACCGGTGAGAAATTTGGTACTT 738

QY 201 GlnLysTyrIleSerGluProValValIlePheLeuHisIleIleIleThrMetThrGlu 220
DB 739 CAGAAATACATACAGAACCTGTTGTCATCTTCTTATATATATATATATATATATATAT 798

QY 221 ValLeuTyrProValTyrValThrLeuArgCysAspSerAlaPheLeuSerGlyValThr 240
DB 799 GTTTTGTATCCAGTTTACGTCACCCATAGGTGTGATCTCTGCTTTTATATATATATATAT 858

QY 241 LeuMetLeuThrCysIleValThrPheLysLeuValSerTyrAlaHisThrSerTyr 260
DB 859 TTGATGCTCCTCCTACCTGTGATTTGGCTAAAGTTGTTTCTTATGCTCATACTACTAT 918

QY 261 AspIleArgSerLeuAlaAsnAlaAlaAspLysAlaAsnProGluValSerTyrVal 280
DB 919 GACATAAGATCCCTAGGCCAATCGAGCTGATAGGCCAATCCCTGAGTCTCTACTACGTT 978

```

```

281 SerLeuLysSerLeuAlaTyrPheMetValAlaProThrLeuCysTyrGlnProSerTyr 300
DB 979 AGCTTGAGAGAGCTTGCGATATTTTCATGTCGCTCCACATTTGTGTATCAGCAAGTAT 1038

QY 301 ProArgSerAlaCysIleArgLysGlyTyrPheValAlaAlaArgGlnPheAlaLysLeuValIle 320
DB 1039 CCACGTTCTGCATGTATACGGAAGGTTGGGTGGCTGCAATTTGCAAACTGGGTCA 1098

QY 321 PheThrGlyPheMetGlyPheIleIleGluGlnTyrIleAsnProIleValArgAsnSer 340
DB 1099 TTCCACGGATTCATGGATTTATATAGAACAAATATATAAATCTTATCTCAGGAAC 1158

QY 341 LysHisProLeuLysGlyAspLeuLeuTyrAlaIleGluArgValLeuLysLeuSerVal 360
DB 1159 AAGCATCTCTTTGAAAGGCGATCTTCTATATGCTATTGAAAGAGTGTGGAAGCTT 1218

QY 361 ProAsnLeuTyrValThrLeuCysMetPheTyrCysPhePheHisLeuThrPheAsnIle 380
DB 1219 CCAAAATTTATATGTGTGGCTCTGCATGTTCTACTGCTTCTCCACCTTTGGTTAA 1278

QY 381 LeuAlaGluLeuLeuCysPheGlyAspArgGluPheTyrLysAspTrpIleAsnAlaLys 400
DB 1279 TTGGCAGAGCTTCTCTGCTTCGGGATCGTGAATTCACAAGATTTGGTGAATGCA 1338

QY 401 SerValGlyAspTyrTrpArgMetTrpAsnMetProValHisLysTrpMetValArgHis 420
DB 1339 AGTGTGGGAGATTTACTGGAGAATGTGAATATGCTGTTTCTATAAATGGATGCTG 1398

QY 421 IleTyrPheProCysLeuArgSerLysIleProLysThrLeuAlaIleIleAlaPhe 440
DB 1399 ATATACTTCCCGTGTTCGCGACAGATACCAAGACACTCGCCATATATCATGTGCT 1458

QY 441 LeuValSerAlaValPheHisGluLeuCysIleAlaValProCysArgLeuPheLysLeu 460
DB 1459 CTAGTCTCTGCAGTCTTTCATGAGCTATGCATCGCAGTTCCTTGTCTCTTCAAG 1518

QY 461 TrpAlaPheLeuGlyIleMetPheGlnValProLeuValPheIleThrAsnTyrLeuGln 480
DB 1519 TGGGCTTTCTTGGGATTTATGTTTCAGGTGCGCTTTGCTTCTATCACAACACTAT 1578

QY 481 GluArgPheGlySerThrValGlyAsnMetIlePheThrPheIlePheCysIlePheGly 500
DB 1579 GAAAGTTTGGCTCAACGGTGGGGAACATGATCTTCTGTTCTCTCTCTCTCTCGA 1638

QY 501 GlnProMetCysValLeuLeuTyrTyrHisAspLeuMetAsnArgLysGlySerMetSer 520
DB 1639 CAACCGATGTGTGCTTCTTCTTATTACCGACCTGATGAACCGAAAGGATCGATG 1698

```

RESULT 2

```

US-10-157-855-1
; Sequence 1, Application US/10157855
; Publication No. US20020170091A1
; GENERAL INFORMATION:
; APPLICANT: Lasser, Michael W.
; APPLICANT: Ruzinsky, Diane W.
; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic
; FILE OF INVENTION: Acid Sequences
; FILE REFERENCE: 16516.158
; CURRENT APPLICATION NUMBER: US/10/157,855
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 09/326,203
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,143
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/108,389
; PRIOR FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1942
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: August 31, 2003, 05:57:06 ; Search time 371 Seconds

(without alignments)
3215.662 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771

Sequence: 1 MAILDSAGVTVTENGGEF.....QPMCVLLYHYHDLNMRKGSMS 520

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1533700 seqs, 1147125425 residues

Total number of hits satisfying chosen parameters: 3067400

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlip
-Q=/cgn2_1/USPTO.spool/US09623514/runat_29082003_152202_190/app_query.fasta_1.711
-DB=PublishedApplications_NA_QFMT-fastap -SUFFIX=rnbp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09623514@cgn2_1.1.333@runat_29082003_152202_190
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA:**
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq**
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq**
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq**
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq**
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq**
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq**
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq**
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq**
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq**
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq**
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq**
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq**
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq**
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq**
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq**
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq**
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq**

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	1904	14	US-10-223-076-2 Sequence 2, Appli

2	2771	100.0	1942	13	US-10-157-855-1
3	2351.5	84.9	1537	14	US-10-223-076-4
4	2309.5	83.3	1512	14	US-10-223-076-8
5	1880	67.8	2090	14	US-10-223-076-10
6	1854	66.9	1964	14	US-10-223-076-14
7	1842.5	66.5	2099	14	US-10-223-076-12
8	1833	66.1	1446	14	US-10-223-076-6
9	1279	46.2	1572	14	US-10-223-076-17
10	1057.5	38.5	1181	14	US-10-223-076-16
11	791.5	28.6	1467	14	US-10-278-733-9
12	791.5	28.6	1467	14	US-10-278-733-3
13	773.5	27.9	1650	14	US-10-273-438-3
14	773.5	27.9	1650	14	US-10-273-438-9
15	773.5	27.9	1766	13	US-10-157-855-15
16	773.5	27.9	1766	13	US-10-157-855-16
17	771.5	27.8	1497	14	US-10-278-733-1
18	771.5	27.8	1497	14	US-10-278-733-12
19	762.5	27.5	1497	14	US-10-278-733-3
20	753.5	27.2	1470	14	US-10-278-733-4
21	751.5	27.1	1470	14	US-10-278-733-7
22	724.5	26.1	1497	14	US-10-278-733-5
23	720.5	26.0	1698	14	US-10-278-733-6
24	709.5	25.6	1895	13	US-10-157-855-14
25	698	25.2	380	9	US-09-770-791-192
26	696.5	25.1	1411	14	US-10-273-438-1
27	652.5	23.5	629	14	US-10-273-438-4
28	652.5	23.5	629	14	US-10-223-076-1
29	573.5	20.7	993	9	US-09-764-853-79
30	535	19.3	1122	11	US-09-774-639-29
31	535	19.3	1122	11	US-09-969-730-58
32	526	19.0	470	9	US-09-770-444-209
33	452.5	16.3	665	14	US-10-066-543-746
34	440	15.9	275	10	US-09-878-574-15694
35	421	15.2	774	10	US-09-764-868-279
36	413.5	14.9	1653	12	US-10-303-664A-11
37	413.5	14.9	4011	12	US-10-303-664A-10
38	407.5	14.7	452	10	US-09-867-701-4664
39	376	13.6	542	10	US-09-764-868-587
40	375	13.5	1607	11	US-09-918-026A-10
41	363	13.1	1569	11	US-09-918-026A-3
42	299	10.8	275	13	US-10-157-855-5
43	285.5	10.7	447	11	US-09-918-995-29256
44	284.5	10.3	253	13	US-10-157-855-7
45	284	10.2	234	13	US-10-157-855-3

ALIGNMENTS

RESULT 1
US-10-223-076-2
; Sequence 2, Application US/10223076
; Publication No. US20030074695A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V
; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: UCAL-105CIP3
; CURRENT APPLICATION NUMBER: US/10/223,076
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

use as 10262

no
Seq
is not
the same
as Seq 2
from

THIS PAGE BLANK (uspto)

Qy 423 FPCURSKIPKTLAIIIAFLVSAVPHCLCIAVPCRLFKLWAFGLMFOVPLV-FTNTYLQE 481
 Db 391 KPMRLRGSSKWMARTGVFLASAFHEYLVSPLRMFLWFTGMAQIPLAWFVGRFQ 450
 Qy 482 RFGSTVGNMIFWFIFCFIGOPMCVLLYYHD 511
 Db 451 NY-----GNAAVW-LSLIIGQPIAVLMYVHD 475

RESULT 15

AAB19742
 ID AAB19742 standard; Protein; 500 AA.

XX AC AAB19742;

XX DT 19-FEB-2001 (first entry)

XX DE Rat acyl CoA:cholesterol acyltransferase.

XX KW Acyl CoA:cholesterol acyltransferase; ACAT; transgenic plant;
 KW sterol; tocopherol; phytosterol; phytostanol; anticholesterolemic;
 KW hypolipemic; rat.

XX OS Rattus sp.

XX PN W0200061771-A2.

XX PD 19-OCT-2000.

XX PF 12-APR-2000; 2000WO-US09696.

XX PR 12-APR-1999; 99US-0128995.

XX PA (MONS) MONSANTO CO.

XX PI Venkatramesh M, Corbin DR, Bhat GB, Boddupalli SS, Grebenok RJ;
 PI Kishore GM, Lardizabal KD, Lassner MW, Rangwala SH, Karunanandaa B;

XX DR WPI; 2000-665136/64.

XX DR N-PSDB; AAA88846.

XX PT Genetically engineering the biosynthetic pathways in plants involved in
 PT the accumulation of sterol compounds and tocopherol to produce
 PT compounds for lowering the level of low density lipoprotein cholesterol
 PT in blood serum

XX PS Disclosure; Page 64; 166pp; English.

XX CC The present sequence is that of rat acyl CoA:cholesterol
 CC acyltransferase (ACAT), as deduced from ACAT DNA (see AAA88846).
 CC Sterol O-acyltransferases such as ACAT catalyse the formation of
 CC cholesterol esters from cholesterol and long chain fatty acids.
 CC Recombinant constructs of the invention are used to alter the
 CC biosynthesis and accumulation of sterols and tocopherols in
 CC transgenic plants. Seeds of such plants may contain elevated
 CC levels of sitostanol and/or its esters, and alpha-tocopherol, and
 CC reduced levels of campesterol and campestanol and their esters.
 CC The seeds may also contain the novel sterol brassicatanol. Oil
 CC obtained from the seeds can be used in food and pharmaceutical
 CC compositions to lower levels of low density lipoprotein cholesterol
 CC in blood serum. ACAT enzymes can be used in the present invention
 CC to produce elevated levels of phytosterol and/or phytostanol esters.

XX SQ Sequence 500 AA;

Query Match 27.9%; Score 773.5; DB 21; Length 500;

Best Local Similarity 37.5%; Pred. No. 1.4e-69;

Matches 194; Conservative 86; Mismatches 167; Indels 71; Gaps 18;

Qy 14 ENGSGEVDLRLRRRSKSDSNGL-LLSGSDNNPSDDVADYDRIDSDVNDDAQ 72

Db 22 QGGSGPMVDEEVR-----DAAVGPDLAGGDAPAPA-PVPAPAHTRDK-----DRQ 67

Qy 73 GTANLAGDNNGGDNNGGGGGREGGRGNADATFTYRPSVPAHRARERPLSSDAIFKQSH 132
 Db 68 TSV-----GDGHWELR-----CH-RLQDSLFSSDSGF-SNY 96
 Qy 133 AGLENLCVVLIIVNSRLIENIMKYGLWINTDFWFSRSLRD---WPLFMCCISLSIFP 189
 Db 97 RGIILNCVVMILILSNARLEENLIKYGILV-DPIQVVSLEFKDPYSWPAPCLIIASNIFI 155
 Qy 190 LAFTVEKLVLQKYISBPVWIFLIIITMTTEVLYPVVTLRCDSAFLSGVTLMLLT-CIV 248
 Db 156 VATFOIEKRUSVGALTEQMGLLHVNLIATIIICFPAAVALLVESITTPVGSFLFALASYII 215
 Qy 249 WLKLVSYAHTSY-----DIRSLANAADK-----ANPEVSY--YVSLKSLAYFWVAPTL 295
 Db 216 FLKLFYRDVNLWCRRVRKAKAVSAGKKVSGAAQNTVSPDNLTYRDLVYFFAPTL 275
 Qy 296 YQPSYRSACIRKGVARQFAKLVFTGFMGFIIEQVINIVRNKHLKGLDYA--IE 353
 Db 276 YELNFRSPRIKRRFLRRVLEMLFTTQLQVGLIQQWVPTIONSMPFK-DMDYSRIE 334
 Qy 354 RVLKLSVNLVWLCMEYCFPHLMLNLAELLCLCFDREFYKDWNAKSVGDYWRWNMPV 413
 Db 335 RLLKLVPNHLIWLFIYFWLFSCLNAVAELLQFGDREFYKDWNAESVTFWQWNIPV 394
 Qy 414 HKWVRHIYFPCLRSKIPKTLAIIIAFLVSAVPHCLCIAVPCRLFKLWAFGLMFOVPLV 473
 Db 395 HKWCIRHFYKPMRLRSGSKWMARTGVFLASAFHEYLVSPLRMFLWAFATMAAQVPLA 454
 Qy 474 FTINYLQERFGSTVGNMIFWFIFCFIGOPMCVLLYYHD 511
 Db 455 WIVNRF---FQGNYGNAAVW-VTLIIGQPVAVLMYVHD 488

Search completed: August 31, 2003, 04:55:00

Job time : 73 secs

XX 02-DEC-1998; 98US-0110602.
 PR 31-MAR-1999; 99US-0127111.
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Cahoon EB, Kinney AJ, Cahoon RE;
 PI
 XX
 XX WPI; 2000-412308/35.
 DR N-PSDB; AAA48936.
 XX
 XX Polynucleotides encoding diacylglycerol acyltransferase, useful for
 PT synthesis of triacylglycerols and increasing the level of oils in plant
 PT seeds
 XX
 XX Claim 12; Page 47-48; 62pp; English.
 XX
 CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,
 CC soybean and wheat were screened for sequences with homology to a
 CC putative acyl CoA cholesterol acyltransferase related gene from
 CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
 CC sapiens and Mus musculus. The cDNA clones identified from this process
 CC were used to form complete diacylglycerol acyltransferase cDNA
 CC sequences. The present sequence is corn diacylglycerol
 CC acyltransferase protein, derived from a contig of clones
 CC p0042.cspaf49r, p0122.ckmab57r and p0125.czaau61rb. Diacylglycerol
 CC acyltransferases are involved in the synthesis of triacylglycerols.
 CC Alteration of the expression of the diacylglycerol acyltransferase
 CC DNA can be useful for increasing the level of oils in plant seeds.
 CC Inhibitors of diacylglycerol acyltransferase may be useful as
 CC herbicides.
 XX
 XX Sequence 285 AA;
 SQ
 Query Match 37.2%; Score 1030.5; DB 21; Length 285;
 Best Local Similarity 68.3%; Pred. No. 4.7e-96;
 Matches 185; Conservative 30; Mismatches 47; Indels 9; Gaps 3;
 QY 253 VSYATSDIRSLANADK-----ANPEVSYVSLKSLAYFMVAPILCYOPSPRSAC 305
 DB 11 VSAHTNDIRVLSKSTEGAGAYGVDPENMKDPTFKSLVYFMPLAPTLCYOPTPTTC 70
 QY 305 IRKGVARQFAKLVFTGFMGFIIOYINPIVRNSKHPKGLDGLLVAIERVLKSLVPLXV 365
 DB 71 IRKGVWTOOLIKCVVETGLMGLFIIOYINPIVRNSKHPKGLNGLNVAIERVLKSLVPLXV 130
 QY 366 WLCMFYCFHMLNLAELLCGDFEYKDMWNKASVGDYWRMNNPVHKWVRHIYFPC 425
 DB 131 WLCMFYCFHMLNLAELLCGDFEYKDMWNKASVGDYWRMNNPVHKWVRHIYFPC 190
 QY 426 LRSKIPKTLAIIAFVLSAVFHELCIAVPCRLFKLWAFGLIMFOVPLVFTITNQLQERGS 485
 DB 191 IRXGSRGVALIISFLVSAVFHELCIAVPCRLFKLWAFGLIMFOVPLVFTITNQLQERGS 250
 QY 486 T-VGNMFIWFIFICQPMCVLLYYHDLNMR 515
 DB 251 VMVGNMFIWF-PSIVRQPMKXLYNKHDMQK 280
 RESULT 14
 AAB15200
 ID AAB15200 standard; protein; 488 AA.
 XX
 AC AAB15200;
 XX
 XX 14-DEC-2000 (first entry)
 DT
 XX Human ACAT Related Gene Product 1 ARGPI.
 DE
 XX Human; ACAT Related Gene Product 1; ARGPI; gene therapy;
 KW enzyme; acyl Coenzyme A-cholesterol acyltransferase 1; ACAT1;
 KW sterol esterification; lipid homeostasis; diacylglycerol acyltransferase;
 KW DGAT; cholesterol; triglyceride biosynthesis; hypertriglyceridaemia;

KW hyperlipidaemia; atherosclerosis; heart disease; obesity.
 XX Homo sapiens.
 XX US6100077-A.
 XX 08-AUG-2000.
 XX 01-OCT-1998; 98US-0165042.
 XX 01-OCT-1998; 98US-0165042.
 XX (UYCO) UNIV COLUMBIA NEW YORK.
 XX Sturley SL, Oelkers P;
 PI WPI; 2000-557622/51.
 DR N-PSDB; AAA76169.
 XX
 XX New nucleic acid encoding a human diacylglycerol acyltransferase,
 PT useful for treating hyperlipidaemia, atherosclerosis, heart disease, or
 PT other diseases associated with an imbalance of triglyceride levels.
 XX
 XX Claim 1; Fig 1A; 32pp; English.
 XX
 CC The enzyme acyl Coenzyme A-cholesterol acyltransferase 1 (ACAT1) mediates
 CC sterol esterification, an important component of intracellular lipid
 CC homeostasis. The present sequence is human ACAT Related Gene Product 1
 CC (ARGPI). This enzyme is a diacylglycerol acyltransferase (DGAT). This
 CC enzyme does not esterify cholesterol. It is thought therefore that ARGPI
 CC participates in the Coenzyme A-dependent acylation of substrate(s) other
 CC than cholesterol e.g. diacylglycerol. Also, ARGPI has a predicted
 CC diacylglycerol binding motif, suggesting that it may perform the last
 CC acylation in triglyceride biosynthesis. ARGPI gene and protein are useful
 CC for treating a subject who has an imbalance in triglyceride levels due to
 CC a defect in esterification of diglycerol, via gene therapy. Particularly,
 CC ARGPI is useful for treating hypertriglyceridaemia, hyperlipidaemia,
 CC atherosclerosis, heart disease, obesity or other diseases associated with
 CC high or excessive levels of triglyceride.
 XX
 XX Sequence 488 AA;
 SQ
 Query Match 28.6%; Score 791.5; DB 21; Length 488;
 Best Local Similarity 38.0%; Pred. No. 2e-71;
 Matches 194; Conservative 77; Mismatches 174; Indels 65; Gaps 17;
 QY 25 RLRRKRSKSDSNGLLSGSDNNPSDDVGPADYVRDRIDSDVNDDAQG-TANLAGDNG 83
 DB 8 RRRRTGSRPSSHG-----GGPAAAE-----VRDAAGPDVGAAGDAPA 48
 QY 84 GDDNNGGCGGEGEGEGRNADATYTPSPVPAHRRARESPSSDAIFKQSHAGLNFNCVVL 143
 DB 49 PAPNKDGDAGVSGSHWELRC-----HRLQDSLFSDDSGF-SNYRGILNMCVNL 96
 QY 144 IAVNSRLIENIMKYGLIRTDWFSSRLRD---WPLFMCCISLIPPLAFTVEKVL 200
 DB 97 ILSNARLENLIKYGLV-DPIQVVSLEKDPHSPAPCLVIAANVFAAFQVEKRLA 155
 QY 201 OKYISEPVVIFLHIITTEVLYPVYVTLRCDSAFSLGVTMLLT-CIWLKLSVYAHTS 259
 DB 156 VGLTEQAGLLHLVLANLATILCFPAVLLVSEIIPVGSLLALMAHTILFLFSYR--- 212
 QY 260 YDIRS-----LANAADKAMP-----EVSY--YVSLKSLAYFVAPILCYOPSPRSA 304
 DB 213 -DVNSWCRRAKAAASAGKASSAAAPHTVSYPDNLTYRDLYFFFAPTLCYELNFRSP 271
 QY 305 CIRKGVARQFAKLVFTGFMGFIIOYINPIVRNSKHPKGLDGLLVA--IERVLKLSVNP 362
 DB 272 RIRKRELLRILEMLFFTLQVGLTQQWNPVPTIQNSMKPFK-DMDSYRIERLLKLAVPN 330
 QY 363 LYVWLCMFYCFHMLNLAELLCGDFEYKDMWNKASVGDYWRMNNPVHKWVRHIY 422
 DB 331 HLILWIFFYFLWFLHSCNLNAVAELMQGDFEYKDMWNNSVTFWQNNWNPVHKWVRHIY 390


```

XX SQ Sequence 361 AA;
Query Match 43.7%; Score 1211; DB 21; Length 361;
Best Local Similarity 61.1%; Pred. No. 2.8e-114;
Matches 218; Conservative 49; Mismatches 80; Indels 10; Gaps 4;

Qy 168 FSSRLRDWPLMCCISLIPPLAFTVEKLVQKYSIEPVVIFLHIITMTVEVLYPVV 227
Db 1 FNATSLRDWPLLMCCLSLIPPLGAFAVEKLAENNLVSDPATTCFHILFTTTEIVYV 60

Qy 228 TLRCDSAFSLGVTMLTLCIVWKLVSVAHTSYDIRSLANAADKANPEVS-----YV 280
Db 61 ILKCDSAVLSGFLVLMFACIIVWKLVSFAHTNDHDKLITSGKKVDNELTAAGIDNLQXP 120

Qy 281 SLKSLAVFMVAPTLCYQPSYPRACIRKGNVAFQAKLVFTCTGFMGFIIEQYINPIVRN 339
Db 121 TLGSLTFKMAPTLCYQAKVILRPYVRKGLVQRVILYLFTGLQGFIIIEQYINPIVRN 180

Qy 340 SKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFHMLNLILAEILCFGDREFFYKDWNA 399
Db 181 SOHPLMGGLLNAVETVLKSLPNVYLWLCMFYCLFHLNLILAEILREGDREFFYKDWNA 240

Qy 400 KSVGDYWRMWNMPVHKWVRHIYPCLRSKIPKTLAIILAFVSAVPHELCIAVPCRLFK 459
Db 241 KTIDEYWRKWNMPVHKWVRHIYPCLRNKISKEVAVFISFFVSATLHEY-VLLFLHLK 299

Qy 460 LWAFGLIMFQVPLVFTINTYQERFGST-VGNMIFWFCIFGQPMCVLLYYHDLNMR 515
Db 300 FWAFGLIMQLPIILILSYLNKFSWDVGNMIFWFFCYIGQPMCVLLYYHDVNNR 356

RESULT 12
AAY94515
ID AAY94515 standard; Protein; 327 AA.
XX AC AAY94515;
XX DT 06-DEC-2000 (first entry)
XX DE Corn diacylglycerol acyltransferase protein #3.
XX KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
XX KW triacylglycerol; herbicide; EC2.3.1.20.
XX OS zea mays.
XX PN WO200032756-A2.
XX PD 08-JUN-2000.
XX PF 01-DEC-1999; 99WO-US28354.
XX PR 02-DEC-1998; 98US-0110602.
XX PR 31-MAR-1999; 99US-0127111.
XX (DUPO) DU PONT DE NEMOURS & CO E I.
XX PA Cahoon EB, Kinney AJ, Cahoon RE;
XX PI WPT; 2000-412308/35.
XX DR N-PSDB; AAA48935.
XX PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for
XX PT synthesis of triacylglycerols and increasing the level of oils in plant
XX PT seeds.
XX PS Claim 12; Fig 1; 62pp; English.
XX CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,
XX CC soybean and wheat were screened for sequences with homology to a
XX CC putative acyl CoA cholesterol acyltransferase related gene from
XX CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo

```

```

CC sapiens and Mus musculus. The cDNA clones identified from this process
CC were used to form complete diacylglycerol acyltransferase cDNA
CC sequences. The present sequence is corn diacylglycerol acyltransferase
CC protein derived from clone cp1c.pk005.h23. Diacylglycerol
CC acyltransferases are involved in the synthesis of triacylglycerols.
CC Alteration of the expression of the diacylglycerol acyltransferase DNA
CC can be useful for increasing the level of oils in plant seeds.
CC Inhibitors of diacylglycerol acyltransferase may be useful as
CC herbicides.
XX SQ Sequence 327 AA;
Query Match 37.8%; Score 1046.5; DB 21; Length 327;
Best Local Similarity 56.0%; Pred. No. 1.3e-97;
Matches 191; Conservative 45; Mismatches 72; Indels 33; Gaps 4;

Qy 166 FWFSSRLRDWPLMCCISLIPPLAFTVEKLVQKYSIEPVVIFLHIITMTVEVLYPV 225
Db 1 FWFNATSLRDWPLLMCCLSLIPPLGAFAVEKLAENNLVSDPATTCFHILFTTTEIVYV 60

Qy 226 YVTLRCDSAFSLGVTMLTLCIVWKLVSVAHTSYDIRSLANAADKANPEVS-----Y 278
Db 61 LVILKCDSAVLSGFLVLMFACIIVWKLVSFAHTNDHDKLITSGKKVDNELTAAGIDNLQ 120

Qy 279 YVSLKSLAYFMVAPTLCYQPSYPRACIRKGNVAFQAKLVFTCTGFMGFIIEQYINPIVR 338
Db 121 APTLGLSTTFKMAPTLCYQPSYPRTPYVRKGLVQRVILYLFTGLQGFIIIEQYINPIVR 180

Qy 339 NSKHPLKGDLLYAIERVLKLSVPNLYVWLCMFYCFHMLNLILAEILCFGDREFFYKDWNN 398
Db 181 NSQHPLMGGLLNAVETVLKSLPNVYLWLCMFYCLFHLNLILAEILREGDREFFYKDWNN 240

Qy 399 AKSVGDYWRMWNMPVHKWVRHIYPCLRSKIPKTLAIILAFVSAVPHELCIAVPCRLF 458
Db 241 AKTIDEYWRKWNMPVHKWVRHIYPCLRNKISKEVAVFISFFVSATLHEY----- 293

Qy 459 KLWAFGLIMFQVPLVFTINTYQERFGSTVGNMIFWFCIF 499
Db 294 -----LLFHSSSAYI-NVI-----VLYFQMCPP 315

RESULT 13
AAY94516
ID AAY94516 standard; Protein; 285 AA.
XX AC AAY94516;
XX DT 06-DEC-2000 (first entry)
XX DE Corn diacylglycerol acyltransferase protein #4.
XX KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
XX KW triacylglycerol; herbicide; EC2.3.1.20.
XX OS zea mays.
XX FH Key Location/Qualifiers
XX FT Misc-difference 148 /note= "encoded by TAA"
XX FT Misc-difference 164 /note= "encoded by GNC"
XX FT Misc-difference 193 /note= "encoded by NAA"
XX FT Misc-difference 269 /note= "encoded by TNG"
XX FT Misc-difference 274 /note= "encoded by TAA"
XX PN WO200032756-A2.
XX PD 08-JUN-2000.
XX PF 01-DEC-1999; 99WO-US28354.

```

XX	06-DEC-2000 (first entry)	
XX	Wheat diacylglycerol acyltransferase protein #2.	
XX	Diacylglycerol acyltransferase; corn; rice; soybean; wheat;	
KW	triacylglycerol; herbicide; EC2.3.1.20.	
XX	Triticum aestivum.	
XX	WO200032756-A2.	
PN	08-JUN-2000.	
PD	01-DEC-1999; 99WO-US28354.	
XX	02-DEC-1998; 98US-0110602.	
XX ²	31-MAR-1999; 99US-0127111.	
PR	(DUPO) DU PONT DE NEMOURS & CO E. I.	
XX	Cahoon EB, Kinney AJ, Cahoon RE;	
XX	WPI; 2000-412308/35.	
DR	N-PSDB; AAA48942.	
DR		
XX	Polynucleotides encoding diacylglycerol acyltransferase, useful for	
PT	synthesis of triacylglycerols and increasing the level of oils in plant	
PT	seeds	
XX	Claim 12; Fig 1; 62pp; English.	
PS		
XX	In the present invention, cDNA libraries from Arabidopsis, corn, rice,	
CC	soybean and wheat were screened for sequences with homology to a	
CC	putative acyl CoA cholesterol acyltransferase related gene from	
CC	Arabidopsis thaliana and diacylglycerol acyltransferases from Homo	
CC	sapiens and Mus musculus. The cDNA clones identified from this process	
CC	were used to form complete diacylglycerol acyltransferase cDNA	
CC	sequences. The present sequence is wheat diacylglycerol	
CC	acyltransferase protein, derived from clone wrl-pk0119.b6:fls.	
CC	Diacylglycerol acyltransferases are involved in the synthesis of	
CC	triacylglycerols. Alteration of the expression of the diacylglycerol	
CC	acyltransferase DNA can be useful for increasing the level of oils in	
CC	plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful	
CC	as herbicides.	
XX		
QQ	Sequence 508 AA;	
	Query Match 57.6%; Score 1596; DB 21; Length 508;	
	Best Local Similarity 68.7%; Pred. No. 3,2e-153;	
	Matches 301; Conservative 44; Mismatches 81; Indels 12; Gaps 4;	
QY	88 NGGGRGCGEGGNADATFTTPSV-PAHRRARESPSSDAIFKQSHAGLNLGVVLIIV 146	
DB	69 HGEAAGAAAAARRDALL---PGVGAARRVKESPLSSDAIFRQSHAGLNLGVVLIIV 125	
QY	147 NSRLIIENLMKYGLWIRTFDFWFSRSRLDWPFLFMCISLFIPLAAFTVEKLVLQKYISE 206	
DB	126 NSRLIIENLMKYGLLIIRAGWFWSARSIGDWPMLLCOLLTFPIPLAALTEKWAQRKLIRD 185	
QY	207 PWVIFLHIITMTVELYVPVYVTRCDSAFISGVITLMLTLCIVWLKLVSYAHTSIDRSLA 266	
DB	186 HVSTLLHIITTTVLIYPPVVVILKCESAVLSGFLMFIASITWLKLVSYFAHTNYDIRLS 245	
QY	267 NAADKA-----NPEYSYVYVLSKSLAYFWVAPTLCPQSPYPRACIRKGVARQFAKLY 319	
DB	246 QSTKEGATHGSSIDEENIKGPTINSVYVYFMLAPTLCPQSPYPRATIRKGVWTRQLKCV 305	
QY	320 IFTCFMGFIIEOYNINPIVRNSKHPKLGDLIYAIERVLKLSYPNLYVWLVCNFCYFHLWLN 379	
DB	306 VFTGLMGFIIEQYNINPIVQNSKHPNLGNFLDAIERVLKLSVPTLYVWLVCNFCYFHLWLN 365	
QY	380 ILAELLFCGDRFYEYKDWNNAKSGVDYWRWMNPNVHKWVRHIYFPCLRSKIPKTLAIIIA 439	

366 ILAELLRFGRFEYKDMWNAKTVEYWRQWNNPVHKWILVRHLYFPCIRNGUSKGCALLIA 425

440 FLNSAVFHELCIAVPCRLFKLWAFIGIMFQVPLVFITNYLQERFGST-VGNNMIFWFICI 498

426 FLNSAVFHELCIAVPCRFKLSWAGIMFQVPLLETKYLDKFTNTVGNMIFWFESI 485

499 FGQPMCVLLYYHDLNMRK 516

486 VGQPMCVLLYYHDMNRQ 503

RESULT 11

AAAY94513

ID AAY94513 standard; Protein; 361 AA.

XX AC AAY94513;

XX 06-DEC-2000 (first entry)

XX Corn diacylglycerol acyltransferase protein #1.

XX Diacylglycerol acyltransferase; corn; rice; soybean; wheat;

XX triacylglycerol; herbicide; EC2.3.1.20.

XX Zea mays.

XX Key Location/Qualifiers

XX Misc-difference 95 /note= "encoded by ATAA"

XX Misc-difference 119 /note= "encoded by NCT"

XX Misc-difference 142 /note= "encoded by CTN"

XX Misc-difference 290..295 /note= "encoded by TTATGTGTGCTGTTCCCTGC"

XX WO200032756-A2.

XX PN 08-JUN-2000.

XX PD 01-DEC-1999; 99WO-US28354.

XX PF 02-DEC-1998; 98US-0110602.

XX PR 31-MAR-1999; 99US-0127111.

XX PA (DUPO) DU PONT DE NEMOURS & CO E. I.

XX PI Cahoon EB, Kinney AJ, Cahoon RE;

XX DR WPI; 2000-412308/35.

XX DR N-PSDB; AAA48933.

XX Polynucleotides encoding diacylglycerol acetyltransferase, useful for

PT synthesis of triacylglycerols and increasing the level of oils in plant

PT seeds

XX Claim 12; Page 42-43; 62pp; English.

XX In the present invention, cDNA libraries from Arabidopsis, corn, rice,

CC soybean and wheat were screened for sequences with homology to a

CC putative acyl CoA cholesterol acyltransferase related gene from

CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo

CC sapiens and Mus musculus, the cDNA clones identified from this process

CC were used to form complete diacylglycerol acyltransferase cDNA

CC sequences. The present sequence is corn diacylglycerol

CC acyltransferase protein, derived from a contig of clones

CC cpjlc.pk005.h23, cen3n.pk010.c10, cco1.pk0029.b6. Diacylglycerol

CC acyltransferases are involved in the synthesis of triacylglycerols.

CC Alteration of the expression of the diacylglycerol acyltransferase

CC DNA can be useful for increasing the level of oils in plant seeds.

CC Inhibitors of diacylglycerol acyltransferase may be useful as

CC herbicides.

SQ	Sequence	504 AA;	
	Query Match	65.7%; Score 1821; DB 21; Length 504;	
	Best Local Similarity	66.5%; Pred. No. 4.2e-176;	
	Matches 353; Conservative 47; Mismatches 91; Indels 40; Gaps 9;		
QY	1 MAILDS-AGVTVTENGGEFVDLRLRRKRSRSSNGLLLSGSDNNSPSDVGCAPADV 59	: : :	
DB	1 MAIDPEPSVATALHSS-----LRRPS-ATSTAGLFNSPETTTDSSGDDDLAKDSG 51	: : :	
QY	60 RDRIDSVDADAGTANLAGDNNGGDNGGGRGGEGGNADAT---FTVRPSVPAHRR 116	: : :	
DB	52 SD--DSINSDDA--AVNSQQONE-----KQDVFSLAFATRPSVPAHRK 92	: : :	
QY	117 ARESPSSDAIFKQSHAGLNLVAVNSRLIENLMKYGLWLIIRDFWSSSLRDW 176	: : :	
DB	93 VKESPLSSDTIFRQSHAGLNLVAVNSRLIENLMKYGLWIKSGFESSSLRDW 152	: : :	
QY	177 PLFMCISLSTPLAFTVEKLVQKYSERPVPFLHIITMTVEVLYPVYVTLRCDAPL 236	: : :	
DB	153 PLFMCCLSLVFPFAFIVEKLAQRKCIPEPVVVLHIITSTLSFLYVPLVILRCDSAFV 212	: : :	
QY	237 SGVTMLTCTIWLKLVSAHTSYDIRSLANAADKA-----NPEVSYYSLSLAYEM 289	: : :	
DB	213 SCVTMLTSCVWLKLVSAHTSYDIRSLANAADKA-----NPEVSYYSLSLAYEM 272	: : :	
QY	290 VAPTLCYQSPRSACIRKGVARQFAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDL 349	: : :	
DB	273 VAPTLCYQSPRTPYIRKGLVQLVLIIFTGVMGFIIDQYINPIVQNSQHPKLGDL 332	: : :	
QY	350 YAIERVLKSLVPLNLYWLCMYCFPHLMNLIAELLCFGRDFEYKDMWNAKSVGDYWRMW 409	: : :	
DB	333 YATERVLKSLVPLNLYWLCMYCFPHLMNLIAELLCFGRDFEYKDMWNAKSVGDYWRMW 392	: : :	
QY	410 NMPVHKWVRHYFPCLRSKIPKTLIIIAFLVSFAVHELCTAVPCRLFKLWAFGIMFO 469	: : :	
DB	393 NMPVHKWVRHYFPCLRHGLPKAAALLIAFLVSFAVHELCTAVPCRLFKLWAFGIMFO 452	: : :	
QY	470 VPLVFITNYLQERF-GSTVGNMIFWIFCIFGQPMCVLLYVHDLNMRKGM 519	: : :	
DB	453 VPLVLTITNYLQERF-GSTVGNMIFWIFCIFGQPMCVLLYVHDLNMRKGM 503	: : :	
RESULT 9			
AA94518	standard; Protein; 500 AA.		
XX	AC	AA94518;	
XX	DT	06-DEC-2000 (first entry)	
XX	DE	Rice diacylglycerol acyltransferase protein #2.	
DE	DE		
KW	KW	Diacylglycerol acyltransferase; corn; rice; soybean; wheat;	
KW	KW	triacylglycerol; herbicide; EC2.3.1.20.	
XX	OS	Oryza sativa.	
XX	PH	Key	Location/Qualifiers
FT	FT	Region	1..69
FT	FT	/note= "The nucleotides encoding this region are not given in AAA48938"	
FT	FT	Region	70..500
FT	FT	/note= "Encoded by nucleotides 15 to 1310 of the cDNA in AAA48938"	
XX	XX	WO200032756-A2.	
XX	PD	08-JUN-2000.	
XX	XX	AA94522	
XX	PF	01-DEC-1999; 99WO-US28354.	
XX	XX	02-DEC-1998; 98US-0110602.	
PR	PR		

PR	31-MAR-1999; 99US-0127111.		
XX	(DUPO) DU PONT DE NEMOURS & CO E I.		
PA	Cahoon EB, Kinney AJ, Cahoon RE;		
PI	WPI; 2000-412308/35.		
XX	N-PSDB; AAA48938.		
DR	Polynucleotides encoding diacylglycerol acetyltransferase, useful for		
DR	synthesis of triacylglycerols and increasing the level of oils in plant		
XX	seeds		
PS	Claim 12; Fig 1; 62pp; English.		
XX	In the present invention, cDNA libraries from Arabidopsis, corn, rice,		
CC	soybean and wheat were screened for sequences with homology to a		
CC	putative acyl CoA cholesterol acyltransferase related gene from		
CC	Arabidopsis thaliana and diacylglycerol acyltransferases from Homo		
CC	sapiens and Mus musculus. The cDNA clones identified from this process		
CC	were used to form complete diacylglycerol acyltransferase cDNA		
CC	sequences. The present sequence is rice diacylglycerol		
CC	acyltransferase protein, derived from clone rl324.pk0034.d8.fis.		
CC	Diacylglycerol acyltransferases are involved in the synthesis of		
CC	triacylglycerols. Alteration of the expression of the diacylglycerol		
CC	acyltransferase DNA can be useful for increasing the level of oils in		
CC	plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful		
CC	as herbicides.		
XX	Sequence 500 AA;		
SQ	Query Match	59.1%; Score 1636.5; DB 21; Length 500;	
	Best Local Similarity	65.6%; Pred. No. 2.4e-157;	
	Matches 292; Conservative 61; Mismatches 79; Indels 13; Gaps 3;		
QY	84 GGDNG-----GGGGGGGGRGNADATFYRVSVPARRARSPSSDAIFKQSHAGLNL 138	: : :	
DB	51 GGDNSGRSVLRPGGGGGGGGDFSAFTFRAAAPVHRKAKESPLSSDAIFKQSHAGLNL 110	: : :	
QY	139 CVVLIIVANSRLIENLMKYGLWLIIRDFWSSRLDWPFLPMCCISLSTFPLAFTVEKL 198	: : :	
DB	111 CIVLVAVNSRLIENLMKYGLLIRAGFWFNDKSLRDWPLLMCCLSLPAFPLGAFVEKL 170	: : :	
QY	199 VLQYISEPVIFLHIITMTTEVLYPVYVTLRCDSAFLSGVTMLLTCTIVWLKLVSAHT 258	: : :	
DB	171 AFNVITDAVATCLHIFLSTTEIVYVPLVKCDSAVLSCGFLLIETIACIWLKLVSAHT 230	: : :	
QY	259 SYDIRSLANAADKANPEVSY-----VSLKSLAYFWVAPTLCYQSPRSACIRKGV 311	: : :	
DB	231 NHDIRQLTMGGKVKVDNELSTVDMNLQPTTLGNLIYFMWAPTLCYQSPRTSCYRKGWL 290	: : :	
QY	312 ARQFAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDLIYATIERVLKSLVPLNLYWLCMFY 371	: : :	
DB	291 IRQIILYLIFTGLQGFIEQYINPIVNSQHPKLGKLLNAVETVLKSLVPLNLYWLCMFY 350	: : :	
QY	372 CFFHLWNLIALLLCFGRDFEYKDMWNAKSVGDYWRMNMMPVHKWVRHYFPCLRSRIP 431	: : :	
DB	351 AFFHLWLSILAEILRFGRDFEYKDMWNAKTIDYWRKWNMPVHKWVRHYFPCCRNGIS 410	: : :	
QY	432 KTLAIIIAFLVSAPHELCTAVPCRLFKLWAFGIMFOVPLVFTINYLOERFSGT-VGNM 490	: : :	
DB	411 KEVAVLISFLVSALVHEICVAVPCRLFKWAFGLIMQLIPLVLTAYLKSFRDPTMVGNM 470	: : :	
QY	491 IFWTFICIFGQPMCVLLYVHDLNMR 515	: : :	
DB	471 IFWFFFCIYGQPMCVLLYVHDMNR 495	: : :	
RESULT 10			
AA94522	standard; Protein; 508 AA.		
XX	AC	AA94522;	

XX DGAT; diacylglycerol acyltransferase; seed oil; fatty acid synthesis;
 KW size; weight; carbon flux; TAGI; insertion mutant.
 XX Arabidopsis thaliana.
 OS
 XX Key Location/Qualifiers
 FH Misc-difference 158...184
 FT /label= insertion
 FT /note= "present due to a 147 bp insertion duplication
 in the genomic DNA, see AAA51485".
 XX
 XX WO200036114-A1.
 XX
 XX 22-JUN-2000.
 PD
 XX 16-DEC-1999; 99WO-CA01202.
 XX
 XX 17-DEC-1998; 98US-0112812.
 XX
 XX (CANA) NAT RES COUNCIL CANADA.
 PA
 XX Zou J, Taylor DC, Wei Y, Jaka CC;
 PI WPI; 2000-431592/37.
 DR N-PSDB; AAA51484, AAA51485.
 DR
 XX New DNA encoding diacylglycerol acyltransferase from Arabidopsis
 PT thaliana for transforming plants and regulating seed oil content, fatty
 PT acid synthesis and seed oil acyl composition in commercial and crop
 PT plants
 XX
 PS Claim 5; Page 81-83; 91pp; English.
 XX
 CC This is the Arabidopsis thaliana ecotype Columbia AS11 insertion mutant
 CC diacylglycerol acyltransferase (DGAT). The AS11 mutant TAG1 allele has a
 CC 147 bp insertion located at the central region of intron 2. The insertion
 CC is a duplication of a segment that is composed of 12 bp from the 3' end
 CC of intron 1, the entire sequence of exon 2 (81 bp) and 54 bp from the
 CC 5' end of intron 2. The DGAT and the insertion mutant (AS11) are useful
 CC for regulating seed oil content, the ratio of diacylglycerol to
 CC triacylglycerol proportions in seed oil, fatty acid synthesis, seed oil
 CC acyl composition, seed size/weight and carbon flux into other seed
 CC components in commercial and crop plants. The natural formation of
 CC triacylglycerols can be modified to increase the yield in commercial
 CC plant oils or modify their composition to achieve specific commercial
 CC improvements of plants and plant products.
 XX
 XX Sequence 547 AA;
 SQ
 Query Match 99.2%; Score 2747.5; DB 21; Length 547;
 Best Local Similarity 95.1%; Pred. No. 2.8e-270;
 Matches 520; Conservative 0; Mismatches 0; Indels 27; Gaps 1;
 1 MAILDSAGTWTVTENGGEFVDLRLRRKRSRSDSSNGLLLSGSDNNSPSDDVGAPADVR 60
 1 MAILDSAGTWTVTENGGEFVDLRLRRKRSRSDSSNGLLLSGSDNNSPSDDVGAPADVR 60
 61 DRIDSVNDDAAGTANLAGDNNGGDNGGGRGGEGRGADATFTYRPSVPAHRRARES 120
 61 DRIDSVNDDAAGTANLAGDNNGGDNGGGRGGEGRGADATFTYRPSVPAHRRARES 120
 121 PLSSDAIFKQ-----SHAGLNLVAVNSRLIIE 153
 121 PLSSDAIFKQSHAGLNLVAVNSRLIIE 180
 154 NLMYGWLIRTFDFSSRLRDWFLMCCISLSIFPLAAFTVEKLVQKYSPEVWIFLH 213
 181 NLMYGWLIRTFDFSSRLRDWFLMCCISLSIFPLAAFTVEKLVQKYSPEVWIFLH 240
 214 IIITWTEVLYPVVYTLRCDSAFLSGVTMLLTCTIWLKLVSYAHTSYDIRSLANAADKAN 273
 241 IIITWTEVLYPVVYTLRCDSAFLSGVTMLLTCTIWLKLVSYAHTSYDIRSLANAADKAN 300

QY 274 PEVSYYVSLKSLAYPMVAPTLCYQPSYPSRACIRKGVARQAKLVIFTGFMGFIIEQYI 333
 DB |||||||
 QY 301 PEVSYYVSLKSLAYPMVAPTLCYQPSYPSRACIRKGVARQAKLVIFTGFMGFIIEQYI 360
 DB |||||||
 QY 334 NPVIRNSKHPLKGDLLYAIERVLKLSVPLNVLWLCMFYCFPHLWLNILAEILLCFGDREFY 393
 DB |||||||
 QY 361 NPVIRNSKHPLKGDLLYAIERVLKLSVPLNVLWLCMFYCFPHLWLNILAEILLCFGDREFY 420
 DB |||||||
 QY 394 KDWNAKSVGVYWRMNMNPFVHKMVRHIYFFCLRSKIPKTLAIITAFVLSAVFHELCTAV 453
 DB |||||||
 QY 421 KDWNAKSVGVYWRMNMNPFVHKMVRHIYFFCLRSKIPKTLAIITAFVLSAVFHELCTAV 480
 DB |||||||
 QY 454 PCRLFKLWAFLGIMFQVPLVFTITNYLQERFSGTVGNMIFWIFCITFGQPMCVLLYYHDLML 513
 DB |||||||
 QY 481 PCRLFKLWAFLGIMFQVPLVFTITNYLQERFSGTVGNMIFWIFCITFGQPMCVLLYYHDLML 540
 DB |||||||
 QY 514 NRKGSMS 520
 DB |||||||
 QY 541 NRKGSMS 547
 DB |||||||
 RESULT 8
 AAY94519
 ID AAY94519 standard; Protein; 504 AA.
 XX
 AC AAY94519;
 DT 06-DEC-2000 (first entry)
 XX
 DE Soybean diacylglycerol acyltransferase protein #1.
 XX
 KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
 KW triacylglycerol; herbicide; EC2.3.1.20.
 OS Glycine max.
 XX
 PN WO200032756-A2.
 XX
 PD 08-JUN-2000.
 PF 01-DEC-1999; 99WO-US28354.
 PR 02-DEC-1998; 98US-0110602.
 PR 31-MAR-1999; 99US-0127111.
 XX
 PA (DUPO) DU PONT DE NEMOURS & CO E I.
 XX
 PI Cahoon EB, Kinney AJ, Cahoon RE;
 XX
 DR WPI; 2000-412308/35.
 DR N-PSDB; AAA48939.
 XX
 PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for
 PT synthesis of triacylglycerols and increasing the level of oils in plant
 PT seeds
 XX
 PS Claim 4; Page 51; 62pp; English.
 XX
 CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,
 CC soybean and wheat were screened for sequences with homology to a
 CC putative acyl CoA cholesterol acyltransferase related gene from
 CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo
 CC sapiens and Mus musculus. The cDNA clones identified from this process
 CC were used to form complete diacylglycerol acyltransferase cDNA
 CC sequences. The present sequence is soybean diacylglycerol
 CC acyltransferase protein, derived from clone srl.pk0098.ab.
 CC Diacylglycerol acyltransferases are involved in the synthesis of
 CC triacylglycerols. Alteration of the expression of the diacylglycerol
 CC acyltransferase DNA can be useful for increasing the level of oils in
 CC plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful
 CC as herbicides.
 XX

QY	Db	Query Match	Best Local Similarity	Matches	520 AA;	Score 2771;	DB 22;	Length 520;	Indels	Gaps
QY	1	MAILDSAGVTVVTENGCGGEFVDLDRLRKRSRSDSSNGLLLSGSDNNSPDDVGGAPADVR	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
Db	1	MAILDSAGVTVVTENGCGGEFVDLDRLRKRSRSDSSNGLLLSGSDNNSPDDVGGAPADVR	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
QY	61	DRIDSVVNDDAQGTANLAGDNGGGRGGEGGRGNADATFTYRPSVPAHRRARES	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
Db	61	DRIDSVVNDDAQGTANLAGDNGGGRGGEGGRGNADATFTYRPSVPAHRRARES	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
QY	121	PLSSDAIFKQSHAGLFNLCVYVVIANSRLIENLMKYIGWLIRDTDFWSSRLSDWPLFM	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
Db	121	PLSSDAIFKQSHAGLFNLCVYVVIANSRLIENLMKYIGWLIRDTDFWSSRLSDWPLFM	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
QY	181	CCISLSIFPLAAFTVEKLVLQKYLISEPVWFLHIIITMTVEVLYPVYVTLRCDSAFLSGVT	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
Db	181	CCISLSIFPLAAFTVEKLVLQKYLISEPVWFLHIIITMTVEVLYPVYVTLRCDSAFLSGVT	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
QY	241	LMLLTICVWLKLSYAHSTDIRSLNAADKANPEVSYVYSLKSLAYFMVAPTLCYOPSY	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
Db	241	LMLLTICVWLKLSYAHSTDIRSLNAADKANPEVSYVYSLKSLAYFMVAPTLCYOPSY	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
QY	301	PRSACIRKGWVARQFAKLVIIFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAERVLKLSV	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
Db	301	PRSACIRKGWVARQFAKLVIIFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAERVLKLSV	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
QY	361	PNLYVWLCMYCFCFHLNLNLAELLCPGDRREFYKDWNAKSVGDIYRWNNMPVHKWVRH	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
Db	361	PNLYVWLCMYCFCFHLNLNLAELLCPGDRREFYKDWNAKSVGDIYRWNNMPVHKWVRH	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
QY	421	IYFPCLRSKIPKTLAIIIAFLVSAVFHELCIATVPCRFLKFWAFGLGIMFQVPLVITNVLQ	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
Db	421	IYFPCLRSKIPKTLAIIIAFLVSAVFHELCIATVPCRFLKFWAFGLGIMFQVPLVITNVLQ	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
QY	481	ERFGSTVGNMIFWFIICFICGPMCVLLYIHDLNMRKGSMS	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;
Db	481	ERFGSTVGNMIFWFIICFICGPMCVLLYIHDLNMRKGSMS	100.0%;	0;	0;	100.0%;	1.1e-272;	0;	0;	0;

RESULT 6	
AAV94512	
ID	RAY94512 standard; Protein; 520 AA.
XX	
XX	AAV94512;
XX	
XX	06-DEC-2000 (first entry)
DT	
DT	
XX	
DE	Arabidopsis diacylglycerol acyltransferase clone arabGF protein.
XX	
XX	Diacylglycerol acyltransferase; corn; rice; soybean; wheat;
KW	triacylglycerol; herbicide; EC2.3.1.20.
KW	

Polynucleotides encoding diacylglycerol acetyltransferase, useful for synthesis of triacylglycerols and increasing the level of oils in plant seeds -

Claim 14; Fig 1; 62pp; English.

In the present invention, cDNA libraries from Arabidopsis, corn, rice, soybean and wheat were screened for sequences with homology to a putative acyl CoA cholesterol acyltransferase related gene from Arabidopsis thaliana and diacylglycerol acyltransferases from Homo sapiens and Mus musculus. The cDNA clones identified from this process were used to form complete diacylglycerol acyltransferase cDNA sequences. The present sequence is Arabidopsis thaliana diacylglycerol acyltransferase protein, derived from clone araebcf. Diacylglycerol acyltransferases are involved in the synthesis of triacylglycerols. Alteration of the expression of the diacylglycerol acyltransferase DNA can be useful for increasing the level of oils in plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful as herbicides.

Sequence 520 AA;

Query Match 99.4%; Score 2753; DB 21; Length 520;
Best-Local Similarity 99.6%; Pred. No. 7.2e-271;
Matches 518; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MAILDSAGVTVTVTEGGGEFVDLDRRRKSRSDSSNGLLSGSDNNSPDDVGAPADVR 60
DB 1 MAILDSAGVTVTVTEGGGEFVDLDRRRKSRSDSSNGLLSGSDNNSPDDVGAPADVR 60
QY 61 DRIDSVVNDDAOCTANLAGDNNGGDNGGGRGGEGRGGNADATFYRPSVPAHRARES 120
DB 61 DRIDSVVNDDAOCTANLAGDNNGGDNGGGRGGEGRGGNADATFYRPSVPAHRARES 120
QY 121 PLSDDAIFKOSHAGLFNLCVVLIIVANSRLIIENLMKYGLIRTDFFWSSRLRDWPLFM 180
DB 121 PLSDDAIFKOSHAGLFNLCVVLIIVANSRLIIENLMKYGLIRTDFFWSSRLRDWPLFM 180
QY 181 CCISLSFPLPAAFTVEKLVQKYSBPVVIFLHIITMTVELYPVVYVTLRCDSAFLSGVT 240
DB 181 CCISLSFPLPAAFTVEKLVQKYSBPVGFIFLHIITMTVELYPVVYVTLRCDSAFLSGVT 240
QY 241 LMLTCTIVWLKLYSAHTSYDIRSLANAADKANPEVSYVYVSLKSLAYFWVAPTCYQPSY 300
DB 241 LMLTCTIVWLKLYSAHTSYDIRSLANAADKANPEVSYVYVSLKSLAYFWVAPTCYQPSY 300
QY 301 PRSACIRKGWVARQFAKLVIFTFGMGFIIIEQYINPIVRNSKHPLKGDLLYAIERVLKLSV 360
DB 301 PRSACIRKGWVARQFAKLVIFTFGMGFIIIEQYINPIVRNSKHPLKGDLLYAIERVLKLSV 360
QY 361 PNLVYVWLCMFYCFEHLWNLNLIAELLFCGDBREFYKDWNNAKSVGDIYWRNMNMPYHKWVRH 420
DB 361 PNLVYVWLCMFYCFEHLWNLNLIAELLFCGDBREFYKDWNNAKSVGDIYWRNMNMPYHKWVRH 420
QY 421 IYFPCLRSKIPKTLAIITIAFLVSAVFHELICIAVPCRLFKLWAEFLGIMFQVPLFITNYLQ 480
DB 421 IYFPCLRSKIPKTLAIITIAFLVSAVFHELICIAVPCRLFKLWAEFLGIMFQVPLFITNYLQ 480
QY 481 ERFGSTVGNMIFWIFCIFGQPMCVLLYYHDLNMRKGSMS 520
DB 481 ERFGSTVGNMIFWIFCIFGQPMCVLLYYHDLNMRKGSMS 520

RESULT 7
AAV96854
ID AAY96854 standard; Protein; 547 AA.
XX
AC AAY96854;
XX
DT 09-OCT-2000 (first entry)
XX
DE A. thaliana AS11 mutant diacylglycerol acyltransferase.

RESULT 7	
AA96854	
ID	AA96854 standard; Protein; 547 AA.
XX	
XX	AA96854;
XX	
XX	
DT	09-OCT-2000 (first entry)
XX	
DE	A. thaliana AS11 mutant diacylglycerol acyltransferase.

XX OS Arabidopsis thaliana.
 XX PN WO9963096-A2.
 XX PD 09-DEC-1999.
 XX PF 04-JUN-1999; 99WO-US12541.
 XX PR 05-JUN-1998; 98US-0088143.
 XX PR 12-NOV-1998; 98US-0108389.
 XX PA (CALJ) CALGENE LLC.
 XX PI Lassner MW, Ruezinsky DM;
 XX DR WPI; 2000-105701/09.
 XX DR N-PSDB; AAZ45371.
 XX PT Novel polynucleotides used for modifying plant oil composition and for
 XX PT developing products for treating e.g. cancer, diabetes, cardiopulmonary
 XX PT disease or metabolic disorders
 XX PS Example 2; Page 65-67; 89pp; English.
 XX CC The present sequence represents an acyl-CoA:cholesterol acyltransferase
 XX CC (ACAT) related protein. The ACAT-like protein is active in the formation
 XX CC of a sterol, ester and/or triacylglycerol from a fatty acyl-CoA and
 XX CC sterol and/or diacylglycerol substrate. The DNA can be used for
 XX CC modifying the lipid composition of plant cells. The ACAT-like protein
 XX CC has diacylglycerol acyltransferase (DAGAT) activity, and so the
 XX CC synthesis of triglycerides can be suppressed or increased using the
 XX CC DNA. The protein can be used to produce plant oils with a modified
 XX CC triglyceride content. The products can also be used to identify
 XX CC antagonists and agonists of DAGAT activity. Such agonists and
 XX CC antagonists are particularly useful in treating or ameliorating
 XX CC diseases associated with DAGAT activity, including diseases associated
 XX CC with altered cellular diacylglycerol concentration or PKC activity,
 XX CC including cancer, diabetes, cardiopulmonary diseases or PKC heart failure,
 XX CC atherosclerosis, adipocytosis, leukaemia, skin carcinoma, fibroblastoma,
 XX CC metabolic disorders, obesity, diseases associated with abnormal lipid
 XX CC metabolism, and diseases associated with abnormal fat absorption,
 XX CC lipoprotein secretion and adipogenesis.
 XX SQ Sequence 520 AA;
 Query Match 100.0%; Score 2771; DB 21; Length 520;
 Best Local Similarity 100.0%; Pred. No. 1.le-272;
 Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MAILDSAGVTVTENGSGEFVLDRLRRKRSRSDSSNGLLLSGSDNNSPDDVGAPADVR 60
 Db 1 MAILDSAGVTVTENGSGEFVLDRLRRKRSRSDSSNGLLLSGSDNNSPDDVGAPADVR 60
 QY 61 DRIDSVNDQAQGTANLAGDNNGGDNNGGGEGGEGGNADATFTYRPSVPAHRRARES 120
 Db 61 DRIDSVNDQAQGTANLAGDNNGGDNNGGGEGGEGGNADATFTYRPSVPAHRRARES 120
 QY 121 PLSSDAIFKQSHAGLFNLCVNVVLIANSRLIENLMKYGLIRDTDFWFSRSLRDWPLFM 180
 Db 121 PLSSDAIFKQSHAGLFNLCVNVVLIANSRLIENLMKYGLIRDTDFWFSRSLRDWPLFM 180
 QY 181 CCISLSIFPLAAFTVEKLVQKYLSEPVIFLHIIITMTVEVLYPVYTLRCDSAFLSGVT 240
 Db 181 CCISLSIFPLAAFTVEKLVQKYLSEPVIFLHIIITMTVEVLYPVYTLRCDSAFLSGVT 240
 QY 241 LMLTCTVWLKLVSYAHTSDIRSLANAADKANPEVSYVSLKSLAFMVAFTLCYOPSY 300
 Db 241 LMLTCTVWLKLVSYAHTSDIRSLANAADKANPEVSYVSLKSLAFMVAFTLCYOPSY 300
 QY 301 PRSACIRKGGWVARQAKLVITFTGPMGTIEQYINPIVRNSKHPKLGDLXAIERVLKLSV 360
 Db 301 PRSACIRKGGWVARQAKLVITFTGPMGTIEQYINPIVRNSKHPKLGDLXAIERVLKLSV 360

QY 361 PNLVYVWLCMEFYCFPHLNLNLIAELLFCGDRFEYKDWNAKSVGDIWRWNNMPVHKWVRH 420
 Db 361 PNLVYVWLCMEFYCFPHLNLNLIAELLFCGDRFEYKDWNAKSVGDIWRWNNMPVHKWVRH 420
 QY 421 IYFPCLSRKIPKTLAIIIAFLVSAVFHELCLIAVPCRLFKLWAFLGIMFQVPLVITNYLQ 480
 Db 421 IYFPCLSRKIPKTLAIIIAFLVSAVFHELCLIAVPCRLFKLWAFLGIMFQVPLVITNYLQ 480
 QY 481 ERFGSTVGNMIFWEIFCIFGQPMCVLLYYHDLNMRKGSMS 520
 Db 481 ERFGSTVGNMIFWEIFCIFGQPMCVLLYYHDLNMRKGSMS 520

RESULT 5

AAU00462
 ID AAU00462 standard; Protein; 520 AA.

AC AAU00462;

DT 31-MAY-2001 (first entry)

DE Arabidopsis thaliana sterol acyltransferase ACAT.

KW Lecithin:cholesterol acyltransferase-like; LCAT; sterol acyltransferase;
 KW acyl CoA:cholesterol acyltransferase-like; ACAT; plant oil; cooking oil;
 KW nutritional supplement; dairy product; food product; salad dressing.

OS Arabidopsis thaliana.

PN W0200116308-A2.

XX 08-MAR-2001.

XX 30-AUG-2000; 2000WO-US33863.

XX 30-AUG-1999; 99US-0152493.

XX (MONS) MONSANTO CO.

PI Lassner M, Van Eenennaam A;

XX WPI; 2001-169010/17.

XX N-PSDB; AAS01106.

PT New isolated nucleic acid encoding plant lecithin:cholesterol
 PT acyltransferase-like or acyl (coenzyme A) CoA:cholesterol
 PT acyltransferase-like polypeptides, for modifying the sterol content and
 PT oil production of plants

PS Example 3; Page 103-104; 127pp; English.

XX The present sequence represents Arabidopsis thaliana
 CC acyl CoA:cholesterol acyltransferase-like (ACAT). Several novel
 CC polynucleotides encoding the plant sterol acyltransferases LCAT
 CC (lecithin:cholesterol acyltransferase-like; AAS01081-AAS01104, AAS01341)
 CC and ACAT (AAS01311-AAS01319) are described. A yeast LCAT related open
 CC reading frame, LROI gene sequence (AAS01342), and a rat ACAT (AAS01105)
 CC cDNA sequence are also described. The polynucleotides encoding LCAT
 CC or ACAT are used to produce LCAT or ACAT polypeptides. They can also
 CC be used in a recombinant construct to transform a host cell (preferably
 CC of a plant) or a plant. The recombinant construct is used to increase or
 CC decrease the sterol content of the host cell or plant. It can be used to
 CC alter oil production of the cell or plant, preferably by increasing it.
 CC The oil of the plant or the plant itself is used as a food product, or
 CC as nutritional or dietary supplements, or in pharmaceutical compositions
 CC for lowering cholesterol. The oil can be used in foods e.g. margarine,
 CC butter, cooking oil, and dressings e.g. salad dressings, mayonnaise,
 CC cheese, processed meat, pasta, sauces, cereals, desserts, dips, chips,
 CC baked goods, pastries, cookies, snack bars, confections, chocolates, and
 CC beverages. The alteration in sterol content and/or composition can also
 CC provide a plant with tolerance to stress and insect damage.

...
PT New DNA encoding diacylglycerol acyltransferase from *Arabidopsis*
PT *thaliana* for transforming plants and regulating seed oil content, fatty
PT acid synthesis and seed oil acyl composition in commercial and crop

Genetically engineering the biosynthetic pathways in plants involved in the accumulation of sterol compounds and tocopherol to produce compounds for lowering the level of low density lipoprotein cholesterol in blood serum -

Disclosure: Page 58-59; 166pp; English.

The present sequence is that of Arabidopsis thaliana acyl CoA:cholesterol acyltransferase (ACAT), as deduced from a full-length ACAT DNA sequence (see AA88835). Sterol O-acyltransferases such as ACAT catalyze the formation of cholesterol esters from cholesterol and long chain fatty acids. Recombinant constructs of the invention are used to alter the biosynthesis and accumulation of sterols and tocopherols in transgenic plants. Seeds of such plants may contain elevated levels of sitosterol and/or its esters, and alpha-tocopherol, and reduced levels of campesterol and campestanol and their esters. The seeds may also contain the novel sterol brassicasterol. Oil obtained from the seeds can be used in food and pharmaceutical compositions to lower levels of low density lipoprotein cholesterol in blood serum. ACAT enzymes can be used in the present invention to produce elevated levels of phytosterol and/or phytosterol esters.

Sequence 520 AA;

```

Query Match      100.0%; Score 2771; DB 21; Length 520;
Best Local Similarity 100.0%; Pred. No. 1.1e-272;
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAILDSAGVTVTENGGEFVDLRLRRKRSRSDSSNGLLSGSDNNSPSDDVGAPADVR 60
DB 1 MAILDSAGVTVTENGGEFVDLRLRRKRSRSDSSNGLLSGSDNNSPSDDVGAPADVR 60
QY 61 DRIDSVNDDAQQGTANLAGDNGGDNNGGGRGEGGEGGNADATFTYRPSVPAHRRARES 120
DB 61 DRIDSVNDDAQQGTANLAGDNGGDNNGGGRGEGGEGGNADATFTYRPSVPAHRRARES 120
QY 121 PLSSDAIFKQSHAGLFNLCVVVLIANSRLIIENLMKYGLIRTDFFWSSRLRDWPLFM 180
DB 121 PLSSDAIFKQSHAGLFNLCVVVLIANSRLIIENLMKYGLIRTDFFWSSRLRDWPLFM 180
QY 181 CCISLSIFPLAAFTVEKLVQKYSVPVIFLHIITMTVEVLYPVYVTLRCDSAFLSGVT 240
DB 181 CCISLSIFPLAAFTVEKLVQKYSVPVIFLHIITMTVEVLYPVYVTLRCDSAFLSGVT 240
QY 241 LMLTCTIVLWKLVSIAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPTICYQPSY 300
DB 241 LMLTCTIVLWKLVSIAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPTICYQPSY 300
QY 301 PRSACIRKGWVARQFAKLVITFTGFMGFIIEQYINPIVRNSKHPKLGDLIIAIEVRLKLSV 360
DB 301 PRSACIRKGWVARQFAKLVITFTGFMGFIIEQYINPIVRNSKHPKLGDLIIAIEVRLKLSV 360
QY 361 PNLVYVLCMYCPFHLMNLAELLCFGDREFYKDWNNAKSVGDYWRWNNPVRHKKWVRH 420
DB 361 PNLVYVLCMYCPFHLMNLAELLCFGDREFYKDWNNAKSVGDYWRWNNPVRHKKWVRH 420
QY 421 IYFPCLSKIPKTLAIIIAELVSAPFHELCTIAPCRFLKFAFLGIMFQVPLFITNLYQ 480
DB 421 IYFPCLSKIPKTLAIIIAELVSAPFHELCTIAPCRFLKFAFLGIMFQVPLFITNLYQ 480
QY 481 ERFGSTVGNMIFWFIIFCIGQPMCVLLYYHDLNLRKGSMS 520
DB 481 ERFGSTVGNMIFWFIIFCIGQPMCVLLYYHDLNLRKGSMS 520

```

RESULT 2

AA94524

ID AA94524 standard; Protein; 520 AA.

XX AA94524;

AC AA94524;

XX 06-DEC-2000 (first entry)

DT

XX DE. A. thaliana diacylglycerol acyltransferase protein variant #1.

XX KW Diacylglycerol acyltransferase; corn; rice; soybean; wheat;

XX KW triacylglycerol; herbicide; EC2.3.1.20.

XX OS Arabidopsis thaliana.

XX PN WO2000032756-A2.

XX PD 08-JUN-2000.

XX PF 01-DEC-1999; 99WO-US28354.

XX PR 02-DEC-1998; 98US-0110602.

XX PR 31-MAR-1999; 99US-0127111.

XX PA (DUPO) DU PONT DE NEMOURS & CO E. I.

XX PI Cahoon EB, Kinney AJ, Cahoon RE;

XX DR WPI; 2000-412308/35.

XX PT Polynucleotides encoding diacylglycerol acetyltransferase, useful for

XX PT synthesis of triacylglycerols and increasing the level of oils in plant seeds

XX PS Example 4; Fig 1; 62pp; English.

XX CC In the present invention, cDNA libraries from Arabidopsis, corn, rice,

XX CC soybean and wheat were screened for sequences with homology to a

XX CC putative acyl CoA cholesterol acyltransferase related gene from

XX CC Arabidopsis thaliana and diacylglycerol acyltransferases from Homo

XX CC sapiens and Mus musculus. The cDNA clones identified from this process

XX CC were used to form complete diacylglycerol acyltransferase cDNA

XX CC sequences. The present sequence is Arabidopsis thaliana diacylglycerol

XX CC acyltransferase protein. This sequence was used for homology

XX CC comparison with the novel diacylglycerol acyltransferases.

XX CC Diacylglycerol acyltransferases are involved in the synthesis of

XX CC triacylglycerols. Alteration of the expression of the diacylglycerol

XX CC acyltransferase DNA can be useful for increasing the level of oils in

XX CC plant seeds. Inhibitors of diacylglycerol acyltransferase may be useful

XX CC as herbicides.

XX SQ Sequence 520 AA;

Query Match 100.0%; Score 2771; DB 21; Length 520;

Best Local Similarity 100.0%; Pred. No. 1.1e-272;

Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MAILDSAGVTVTENGGEFVDLRLRRKRSRSDSSNGLLSGSDNNSPSDDVGAPADVR 60
DB 1 MAILDSAGVTVTENGGEFVDLRLRRKRSRSDSSNGLLSGSDNNSPSDDVGAPADVR 60
QY 61 DRIDSVNDDAQQGTANLAGDNGGDNNGGGRGEGGEGGNADATFTYRPSVPAHRRARES 120
DB 61 DRIDSVNDDAQQGTANLAGDNGGDNNGGGRGEGGEGGNADATFTYRPSVPAHRRARES 120
QY 121 PLSSDAIFKQSHAGLFNLCVVVLIANSRLIIENLMKYGLIRTDFFWSSRLRDWPLFM 180
DB 121 PLSSDAIFKQSHAGLFNLCVVVLIANSRLIIENLMKYGLIRTDFFWSSRLRDWPLFM 180
QY 181 CCISLSIFPLAAFTVEKLVQKYSVPVIFLHIITMTVEVLYPVYVTLRCDSAFLSGVT 240
DB 181 CCISLSIFPLAAFTVEKLVQKYSVPVIFLHIITMTVEVLYPVYVTLRCDSAFLSGVT 240
QY 241 LMLTCTIVLWKLVSIAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPTICYQPSY 300
DB 241 LMLTCTIVLWKLVSIAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPTICYQPSY 300
QY 301 PRSACIRKGWVARQFAKLVITFTGFMGFIIEQYINPIVRNSKHPKLGDLIIAIEVRLKLSV 360
DB 301 PRSACIRKGWVARQFAKLVITFTGFMGFIIEQYINPIVRNSKHPKLGDLIIAIEVRLKLSV 360

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 30, 2003, 19:48:34 ; Search time 71 Seconds

(without alignments)
1162.504 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771

Sequence: 1 MAILDSAGVTVTWGGGEF.....QPMCVLLYYHDLNRKSGMS 520

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneset_19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	520	21	Arabidopsis acyl C
2	2771	100.0	520	21	A. thaliana diacyl
3	2771	100.0	520	21	AA194524
4	2771	100.0	520	21	AA194515
5	2771	100.0	520	22	AA194516
6	2753	99.4	520	22	AAU00462
7	2747.5	99.2	547	21	AA194512
8	1821	65.7	504	21	AA194519
9	1636.5	59.1	500	21	AA194518

10	1596	57.6	508	21	AA194522	Wheat diacylglycer
11	1211	43.7	361	21	AA194513	Corn diacylglycer
12	1046.5	37.8	327	21	AA194515	Corn diacylglycer
13	1030.5	37.2	285	21	AA194516	Corn diacylglycer
14	791.5	28.6	488	21	AA15200	Human ACAT Related
15	773.5	27.9	500	21	AA19742	Rat acyl CoA:chole
16	773.5	27.9	500	21	AA194513	Acyl-CoA:cholester
17	770.5	27.8	497	21	AA194523	Mouse diacylglycer
18	770.5	27.8	489	21	AA194523	Mouse diacylglycer
19	760.5	27.4	489	23	AAE24973	Bovine DGAT1 prote
20	758.5	27.4	467	23	AAE24974	Bovine diacylglyce
21	758.5	27.4	489	24	ABP96046	Bovine DGAT protei
22	756.5	27.3	489	23	AAE24979	Bovine diacylglyce
23	754.5	27.2	489	24	ABP96045	Bovine DGAT protei
24	738.5	26.7	407	19	AAW43406	Human acylcoenzyme
25	724.5	26.1	496	21	AA19741	Caenorhabditis ele
26	724.5	26.1	496	21	AA19741	Acyl-CoA:cholester
27	696.5	25.1	386	21	AA194561	Human diacylglycer
28	612	22.1	155	21	AA194514	Corn diacylglycer
29	573.5	20.7	236	22	ABB10191	Human cDNA SEQ ID
30	573.5	20.7	236	23	ABP66778	Human polyptide
31	526.5	19.0	219	19	AAW43412	Human acylcoenzyme
32	397	14.3	550	19	AAW43409	Human acylcoenzyme
33	396	14.3	550	15	AAW3079	Acetyl coenzyme A:
34	396	14.3	550	19	AAW38416	Human acyl-coenzym
35	378.5	13.7	642	19	AAW43411	Yeast acylcoenzyme
36	375.5	13.6	642	19	AAW38418	Yeast acyl-coenzym
37	375	13.5	525	21	AA197952	Mouse acyl CoA:cho
38	374	13.5	559	22	ABW63696	Drosophila melanog
39	357.5	12.9	522	21	AAW21640	Human ACAT Related
40	351	12.7	502	21	AA197953	Human acyl CoA:cho
41	340.5	12.3	165	22	AAU17635	Novel signal trans
42	307.5	11.1	80	21	AA194517	Rice diacylglycer
43	301	10.9	610	19	AAW43410	Yeast acylcoenzyme
44	301	10.9	610	22	AAW97263	Yeast ARE1 amino a
45	299	10.8	610	19	AAW38417	Yeast acyl-coenzym

ALIGNMENTS

RESULT 1
AA191740
ID AA191740 standard; Protein; 520 AA.
XX
AC AA19740;
XX
XX
DT 19-FEB-2001 (first entry)
XX
DE Arabidopsis acyl CoA:cholesterol acyltransferase.
XX
KW Acyl CoA:cholesterol acyltransferase; ACAT; transgenic plant;
KW sterol; tocopherol; phytosterol; phytostanol; anticholesterolemic;
KW hypolipemic.
XX
OS Arabidopsis thaliana.
XX
PN WO2000061771-A2.
XX
PD 19-OCT-2000.
XX
PF 12-APR-2000; 2000WO-US09696.
XX
PR 12-APR-1999; 99US-0128995.
XX
PA (MONS) MONSANTO CO.
XX
PI Venkatramesh M, Corbin DR, Bhat GB, Boddupalli SS, Grebenok RJ;
PI Kishore GM, Lardizabal KD, Lassner MW, Rangwala SH, Karunanandaa B;
XX
DR WPI: 2000-665136/64.
DR N-PSDB; AAA88835.
XX

ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 STATE STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/121.396
FILING DATE: 07/31/95

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/509.187
FILING DATE: 07/31/95

ATTORNEY/AGENT INFORMATION:
NAME: LAMFORT HAMMITTE, ANN.
REGISTRATION NUMBER: 34,858
REFERENCE/DOCKET NUMBER: DCI-033CPDV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-2700
TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 550 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-121-396-4

Query Match 14.3%; Score 396; DB 2; Length 550;

Best Local Similarity 26.7%; Pred. No. 1.5e-30;

Matches 144; Conservative 75; Mismatches 177; Indels 144; Gaps 24;

```

QY 17 GGEFVDLRLRRRSRDSNGL-----LLSGDNNSPDDVGAPADVDRIDSVVND 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 GSHFDDFVNLIERASLDNGGCAITTSVLEGEKNHNRKDLRAPPE----- 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 71 AQGTANLAGDNGGDDNGGGRGEGGRGNADATFTYPSVPAHRRARESPSSDAIFKQ 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 -QGKIFI-----ARRSLDELLEVDHIRT 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 131 SHAGLNLGVVLIATV-----NSRLITE-NLMKYGW-LIRTDWFSSRLRDWPLFMC 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142 YHMFIALILFILSTLVVDYIDEGRLVLEFSLLSYAFGKFTTVV-----TWIMF 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 183 ISLSIFPLAAFTVEKLVQKY-----ISEPVI-----FLHIITTEVLY-PVYVTLR 230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 -----LSTSVPYFLQHWRTGYKSSHPLIRSLFHGFLFMIFQIGVLGPGTYVVL- 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 231 CDSAPLSGVTMLTLCIVWLKLVSYAHT---SYDIRSLANAADKAN-----PEVSYVSLK 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 ---AYTLPPASRFTIIFQIRFVMAKSHFVRENVPRLNSAKESSTVPIPTVNYL- 298
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 284 SLAFVMAPTLCYOPSPRSACIRKGWVAROFAKLVITFGMGFTIEQYINPIVNSK-H 342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 ---YFLFAPTLIYKDSYPRNTPRWGYVAMFAQVFGCFYVYIFERLCAPLFNRIKOE 355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 343 PLKGDLLYAIERIKLSV-----PNLYVWLCMFCFFHLNLAELICFGDRFYKDW 397
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 PFSA-----RVLVLCVNSILPGVLLILFTFFAFLHCLWNAFAMLRFGDRFYKDW 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 398 NAKSVGDYWRMNNPVMKWHRYH---FPCURSKIPKTLAIIAFLVSAVHELCIAYPC 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 NSTSYSNYRTWNVVVDWLYYAYKDFLWPFSSRFRKSAAMLAVFAVSAVVEHALAV-C 467
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 456 R-----LEKLAFLGIMFQVPLFITNYLQERGSTVGNMIFW-----FFCTFGQ 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 LSFFYPVLFLVLFMFFGMAFN-----FIVNDSRKK---PIWNVLMWTSLSLGLNGVLLCFYSQ 520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 13

PCT-US93-09704A-4

; Sequence 4, Application PC/TUS9309704A
; GENERAL INFORMATION:

; APPLICANT:
; TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
; TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
; NUMBER OF SEQUENCES: 9
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/09704A
; FILING DATE: October 12, 1993
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.S. SER. NO. 959,950
; FILING DATE: October 14, 1992
; APPLICATION NUMBER: U.S. SER. NO. 121,057
; FILING DATE: September 10, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: LAMFORT HAMMITTE, ANN.
; REGISTRATION NUMBER: 34,858
; REFERENCE/DOCKET NUMBER: DCI-033CP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-2700
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 550 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-09704A-4

Query Match 14.3%; Score 396; DB 5; Length 550;

Best Local Similarity 26.7%; Pred. No. 1.5e-30;

Matches 144; Conservative 75; Mismatches 177; Indels 144; Gaps 24;

```

QY 17 GGEFVDLRLRRRSRDSNGL-----LLSGDNNSPDDVGAPADVDRIDSVVND 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 70 GSHFDDFVNLIERASLDNGGCAITTSVLEGEKNHNRKDLRAPPE----- 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 71 AQGTANLAGDNGGDDNGGGRGEGGRGNADATFTYPSVPAHRRARESPSSDAIFKQ 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 -QGKIFI-----ARRSLDELLEVDHIRT 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 131 SHAGLNLGVVLIATV-----NSRLITE-NLMKYGW-LIRTDWFSSRLRDWPLFMC 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 142 YHMFIALILFILSTLVVDYIDEGRLVLEFSLLSYAFGKFTTVV-----TWIMF 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 183 ISLSIFPLAAFTVEKLVQKY-----ISEPVI-----FLHIITTEVLY-PVYVTLR 230
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 193 -----LSTSVPYFLQHWRTGYKSSHPLIRSLFHGFLFMIFQIGVLGPGTYVVL- 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 231 CDSAPLSGVTMLTLCIVWLKLVSYAHT---SYDIRSLANAADKAN-----PEVSYVSLK 283
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 245 ---AYTLPPASRFTIIFQIRFVMAKSHFVRENVPRLNSAKESSTVPIPTVNYL- 298
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 284 SLAFVMAPTLCYOPSPRSACIRKGWVAROFAKLVITFGMGFTIEQYINPIVNSK-H 342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 ---YFLFAPTLIYKDSYPRNTPRWGYVAMFAQVFGCFYVYIFERLCAPLFNRIKOE 355
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 343 PLKGDLLYAIERIKLSV-----PNLYVWLCMFCFFHLNLAELICFGDRFYKDW 397
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 PFSA-----RVLVLCVNSILPGVLLILFTFFAFLHCLWNAFAMLRFGDRFYKDW 408
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 398 NAKSVGDYWRMNNPVMKWHRYH---FPCURSKIPKTLAIIAFLVSAVHELCIAYPC 455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 NSTSYSNYRTWNVVVDWLYYAYKDFLWPFSSRFRKSAAMLAVFAVSAVVEHALAV-C 467
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-2700
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 550 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-121-057-4

Query Match 14.3%; Score 396; DB 1; Length 550;
Best Local Similarity 26.7%; Pred. No. 1.5e-30;
Matches 144; Conservative 75; Mismatches 177; Indels 144; Gaps 24;

QY 17 GGEFVLDLRLRRKRSDSNGL-----LLSGDNNSPDDVGPADVRDRIDSVVND 70
DB 70 GSHFDDFVNLIKESASLNGGCAITTFVSLEGEKNNHRAKDLRAPPE----- 117
QY 71 AQTANLAGDNGGGGNGGGRGGGGRGNADATFTYPSVPAHRRARESPSSDAIFKQ 130
DB 118 -QCKIFI-----ARRSLDELLEVDHRTI 141
QY 131 SHAGLENLGVVLIIV-----NSRLIE-NLMKYGW-LIRTDWFSSRLRDPWLFMC 182
DB 142 YHMFIALILFILSTLVVDYIDEGRLVLEFSLSYAFGKPTVVW-----TWIMF 192
QY 183 ISLSIFPLAAFTVEKVLQKY-----ISEPVVI-----FLHIIITMTEVLY-PVYVTLR 230
DB 193 -----LSTFSVPYFLQHWRTGYSKSHPLRSFLHGFLEFQIGVLGFGPTVVV- 244
QY 231 CDSAFSLGVTLMLTICVWLKLVSYAHT---SYDIRSLANAADKAN-----PEVSYVSLK 283
DB 245 ---AYLTPASRIIIFEQIRFVMAHSEFVRENVPVLNSAKEKSTVPTVQVYL--- 298
QY 284 SLAYFVAPTLCTQPSYPSRACIRKGWAKQAKLVITGFMGFIIEQYINPIVRNSK-H 342
DB 299 ---YFLFAPTLIYRDSYPRNPVTWGYVAMKFAQVFCGCFYVYIIFERLCAPLFRNIKE 355
QY 343 PLKGDLLYAIERVLKLV-----PNLYVWLCMFYCFHFLWNLTLAELLFCGDRFYKDW 397
DB 356 PFSA-----RVLVLCVFNILPGVLLILFTFAFLHCLWNAFAEMLRFGDRMYKDW 408
QY 398 NAKSVGDYWRMNMVHKWVRHIY--FPCLSRKIPKTLAIIITAFVLSAVFHELCTAVPC 455
DB 409 NSTSYNYRTNMVNVVHDWLYIYAYKDFLWFFSKRFSKAAVLAVFAVAVVHEYALAV-C 467
QY 456 R-----LFLKMAFLGIMFQVPLVITNYLQERFGSTVGNMIFW-----FIFCIGQ 501
DB 468 LSFFYPVLVLFMFFGMAFN-----FIVNDSRKK---PIWNLMTSLFLGNGVLLCFYSQ 520

RESULT 11
US-08-509-187D-4
Sequence 4, Application US/08509187D
Patent No. 5834283
GENERAL INFORMATION:
APPLICANT: Chang, Ta-Yuan and Chang, Catherine C.Y.
TITLE OF INVENTION: ACYL Coenzyme A: Cholesterol Acyltransferase
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD, LLP
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/509,187D
FILING DATE: 31-JUL-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Lampert Hammitte, Ann
REGISTRATION NUMBER: 34,858
REFERENCE/DOCKET NUMBER: DCI-033cpdv
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 742-4214
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 550 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-509-187D-4

Query Match 14.3%; Score 396; DB 2; Length 550;
Best Local Similarity 26.7%; Pred. No. 1.5e-30;
Matches 144; Conservative 75; Mismatches 177; Indels 144; Gaps 24;

QY 17 GGEFVLDLRLRRKRSDSNGL-----LLSGDNNSPDDVGPADVRDRIDSVVND 70
DB 70 GSHFDDFVNLIKESASLNGGCAITTFVSLEGEKNNHRAKDLRAPPE----- 117
QY 71 AQTANLAGDNGGGGNGGGRGGGGRGNADATFTYPSVPAHRRARESPSSDAIFKQ 130
DB 118 -QCKIFI-----ARRSLDELLEVDHRTI 141
QY 131 SHAGLENLGVVLIIV-----NSRLIE-NLMKYGW-LIRTDWFSSRLRDPWLFMC 182
DB 142 YHMFIALILFILSTLVVDYIDEGRLVLEFSLSYAFGKPTVVW-----TWIMF 192
QY 183 ISLSIFPLAAFTVEKVLQKY-----ISEPVVI-----FLHIIITMTEVLY-PVYVTLR 230
DB 193 -----LSTFSVPYFLQHWRTGYSKSHPLRSFLHGFLEFQIGVLGFGPTVVV- 244
QY 231 CDSAFSLGVTLMLTICVWLKLVSYAHT---SYDIRSLANAADKAN-----PEVSYVSLK 283
DB 245 ---AYLTPASRIIIFEQIRFVMAHSEFVRENVPVLNSAKEKSTVPTVQVYL--- 298
QY 284 SLAYFVAPTLCTQPSYPSRACIRKGWAKQAKLVITGFMGFIIEQYINPIVRNSK-H 342
DB 299 ---YFLFAPTLIYRDSYPRNPVTWGYVAMKFAQVFCGCFYVYIIFERLCAPLFRNIKE 355
QY 343 PLKGDLLYAIERVLKLV-----PNLYVWLCMFYCFHFLWNLTLAELLFCGDRFYKDW 397
DB 356 PFSA-----RVLVLCVFNILPGVLLILFTFAFLHCLWNAFAEMLRFGDRMYKDW 408
QY 398 NAKSVGDYWRMNMVHKWVRHIY--FPCLSRKIPKTLAIIITAFVLSAVFHELCTAVPC 455
DB 409 NSTSYNYRTNMVNVVHDWLYIYAYKDFLWFFSKRFSKAAVLAVFAVAVVHEYALAV-C 467
QY 456 R-----LFLKMAFLGIMFQVPLVITNYLQERFGSTVGNMIFW-----FIFCIGQ 501
DB 468 LSFFYPVLVLFMFFGMAFN-----FIVNDSRKK---PIWNLMTSLFLGNGVLLCFYSQ 520

RESULT 12
US-09-121-396-4
Sequence 4, Application US/09121396
Patent No. 5968749
GENERAL INFORMATION:
APPLICANT: CHANG, TA-YUAN
TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:

ATTORNEY/AGENT INFORMATION:
NAME: Field, Bret E
REGISTRATION NUMBER: 37,620
REFERENCE/DOCKET NUMBER: 6510-105p
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650 327 3400
TELEFAX: 650 327 3231
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 386 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-103-754A-4

Query Match 25.1%; Score 696.5; DB 4; Length 386;
Best Local Similarity 40.9%; Pred. No. 1.7e-60;
Matches 157; Conservative 68; Mismatches 126; Indels 33; Gaps 12;
QY 150 LIENLMRYGWLIRDFWSSRSRD---WPLFMCCISLISIPFLAAFTVEKLVLOKYE 206
Db 1 LELENLIYGLV-DPIQVSLFLKDPYSWAPCLVIAANFVAQVEKRLAVGALTE 59
QY 207 PVVFLHIIITWTEVLYPVVYTLRCDSAFLSGVTMLLT-CIVWLKLVSYAHTSYDIRS- 264
Db 60 QAGLLHLVANLATILCPFAAVVLLVESITPVGSLALMAHTILFLKLSYR---DVNSW 115
QY 265 -----LANAADRANP-----EVSY--YVLSKSLAYFMVAPTLCYQSPRSACIRKGW 310
Db 116 CRRARAKASAGKASSVAAPHTSYDNLTYRDLVYFLFAPTLCYELNFRSPRIRKRF 175
QY 311 VARQAKLVITGFMGTIEQYINPVRNSKHPLKGLLYA--IERVLKLSVPNLYVWLC 368
Db 176 LLRLLEMLFTQLQVGLIQWVPTIONSMMKPKF-DMYSRIERLLKLVNHLIWI 234
QY 369 MFYCFHMLNLAELLFCGDRFEYKQWNNKSGDYWRMNMVHKWVRHIYPPCLRS 428
Db 235 FFYWLFSCLNAVALMQGDRFEYRDMWNSVTFYFQNNWNPVHKWCIRHFYKPMRLR 294
QY 429 KIPKTLAIIIFLVSAPFHELCLIAVPCRLFLKLAFLGIMFQVPLV-FITNLOERFGSTV 487
Db 295 GSSKMWARTGVLASAFFHEHLYSVPLRFLWAFMGMAQIPLAWFVGRFFQNGY---- 350
QY 488 GNMFWFIFCFIGQPMCVLLYYHD 511
Db 351 GNAAVW-LSLIIGQPIAVLMTYVD 373

RESULT 9
US-09-326-203A-23
Sequence 23, Application US/09326203A
Patent No. 644876
GENERAL INFORMATION:
APPLICANT: Lassar, Mike
TITLE OF INVENTION: Ruzinsky, Diane
TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic
FILE REFERENCE: 17045/00/WO
CURRENT APPLICATION NUMBER: US/09/326,203A
CURRENT FILING DATE: 1999-06-04
PRIOR APPLICATION NUMBER: 60/088,143
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/108,389
PRIOR FILING DATE: 1998-11-12
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 23
LENGTH: 409
TYPE: PRT
ORGANISM: murine
PUBLICATION INFORMATION:

JOURNAL: J. Biol. Chem.
VOLUME: 270
PAGES: 26192-26201
DATE: 1995
US-09-326-203A-23
Query Match 14.9%; Score 413; DB 4; Length 409;
Best Local Similarity 29.4%; Pred. No. 2e-32;
Matches 129; Conservative 73; Mismatches 139; Indels 98; Gaps 23;
QY 118 RESPLSSDAIFKQSHAG-----LFNLCVVVLAVN-SRLIIE-NLMKYGW-LIR 163
Db 1 ROSLL--DELFEVDHRTIYHMFIALLLILFVLSITVDYIDEGRLVLEFNLLAYAGKFP 58
QY 164 TDFWSSRSRLRDWPLFMCCISLISIPFLAAFTVEKLVLOKYE-----SEPVIIE-LHII 216
Db 59 TVIW-----TWAMFLSTLSIPYF-----LFQWAGYKSKSHPLIYSLVHGLL 102
QY 217 TWTEVL-----YPVYVTLRCDSAFLSGVTMLLT-CIVWLKLVSYAHT--SYDIRSLANNA 269
Db 103 FLVFGVLGVFPVTVWL---AYLPPASRFLILEQIRLIMKAHSFVRENIPLVNA 158
QY 270 DKAN-----PEVSYVYVLSKSLAYFMVAPTLCYQSPRSACIRKGWVARQAKLVITGTF 324
Db 159 KEKSKDPLPTVNOYL-----YFLFAPTLYRDYPRTPVTRWGYVAMQFLQVFGCLFY 212
QY 325 MGFIIEQYINPVRNSK-HPLKGLLYAIAERVLKLSV-----PNLYVWLCHMFCFPHLWL 378
Db 213 VYIYFELCAPLFRNIKQEPFSA-----RVLVLCVNSILPGVLIILFSEFAFLHCLWL 265
QY 379 NILAELLFCGDRFEYKQWNNKSGDYWRMNMVHKWVRHIYPPCL--RSKIPKTLAI 436
Db 266 NAFAMLRGDRMFKYKQWNNSTSYNYRTNVVVDWLYVYVYKDLLWFFSKRFKSAAM 325
QY 437 IIAFLVSAPFHELCLIAVPCR-----LFKLAFLGIMFQVPLVFTNLOERFGSTVGNM 490
Db 326 LAVFALSAPVHEALAI-CLSYFYPVLVFLVFMFGMAFN----FIVNDSRKR---PIWNI 377
QY 491 IFW-----FIFCFIQG 501
Db 378 MWASLFLGYGLILCFYSQ 396

RESULT 10
US-08-121-057-4
Sequence 4, Application US/08121057
Patent No. 5484727
GENERAL INFORMATION:
APPLICANT: CHANG, TA-YUAN
TITLE OF INVENTION: CHANG, CATHERINE C. Y.
TITLE OF INVENTION: ACYL COENZYME A: CHOLESTEROL
TITLE OF INVENTION: ACYLTRANSFERASE (ACAT)
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 60 STATE STREET, SUITE 510
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/121.057
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: LAMPORT HAMMITTE, ANN.
REGISTRATION NUMBER: 34,858
REFERENCE/DOCKET NUMBER: DCI-033CP

Query Match 28.6%; Score 791.5; DB 3; Length 488;
Best Local Similarity 38.0%; Pred. No. 9.7e-70;
Matches 194; Conservative 77; Mismatches 174; Indels 65; Gaps 17;

QY 25 RLRRKRSRSDSSGNGLLSGSDNNSPSDDVGAPADVRDRIDSVVNDDAQ-TANLAGDNG 83
DB 8 RRRRTGSPSHGG-----GGPAAAE-----VRDAAAGPDVGAAGDAPA 48
QY 84 GGDNNGGGRGGEGGRGNADATFTYRPSVAHRRARESPSLSDAIFKSHAGLFLNCVVL 143
DB 49 PAPKDGAGVSGHWELRC-----HRLQDSLFSDDSGF-SNYRGLNWCVVML 96
QY 144 IAVNSRLIENIMKYGLIRDFWFSRSLRD----WPLFMCCLSLSTIPPLAFTVEKLV 200
DB 97 ILSNARLFLENLIKYGILV-DPIQVWSLFLKDPHWPAPCLVIAANFAVAAFQVKRLA 155
QY 201 QKYLSEPVVFLHIIITMTVEVLVYVVTLCRDSAFSLGVTMLLT-CIVWLKLYSYAHTS 259
DB 156 VGLTEQAGLLHVNATILCFPAAVVLLVESITPVGSLALMAHTILFLKFSYR--- 212
QY 260 YDIRS-----LANAADRANP-----EVSYS--YVSLKSLAYFWAPTLCYQPSYPSA 304
DB 213 -DVNSWCRARAKAASAGKAKASSAAAPHTVSPDNLYRDLVYFLPAPTLCYELNFRSP 271
QY 305 CIRKGVARQAKLIVFTGFMGFIIEQYINPVRNSKHPKGLDLYA--IERVLKLSVNP 362
DB 272 RIRKRELLRLLELFTQLQVGLIQQWVPTIQNSMKPKF-DMDYSRIIERLKLVNP 330
QY 363 LYVWLCMFCFFHLWNLIAELLCFGDREFFYKDWNAKSGVDYWRMNMVPHKMWVRIY 422
DB 331 HLWILFIFFWLFHSCINAVAEELQDREFFYKDWNAKSGVDYWRMNMVPHKMWVRIY 390
QY 423 FPLRSKIPKTLAIIIAFLVSAVPHELCIAVPCRLFLKWLAFGLIMFOVPLV-FITNLOE 481
DB 391 KPLRGSSKWMARTGVFLASAPHEVLSVPLRMFLMARTGMMAQIPLAWFVGRFQG 450
QY 482 RFGTGVNMIFWTFICIFGQPMCVLLYYHD 511
DB 451 NY-----GNAAVW-LSLIIGOPVAVLVMVHD 475

RESULT 5
US-09-326-203A-17
; Sequence 17, Application US/09326203A
; Patent No. 644876
; GENERAL INFORMATION:
; APPLICANT: Lassner, Mike
; APPLICANT: Ruzinsky, Diane
; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic
; FILE REFERENCE: 17045/00/WO
; CURRENT APPLICATION NUMBER: US/09/326.203A
; CURRENT FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,143
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/108,389
; PRIOR FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 500
; TYPE: PRT
; ORGANISM: Rattus sp.
US-09-326-203A-17

Query Match 27.9%; Score 773.5; DB 4; Length 500;
Best Local Similarity 37.5%; Pred. No. 6.1e-68;
Matches 194; Conservative 86; Mismatches 167; Indels 71; Gaps 18;

QY 14 ENGGEFVLDRLRRRSRSDSSNGL-LLSGSDNNSDDVGAPADVRDRIDSVVNDDAQ 72
DB 22 QGSGPMVDEEVR-----DAAVGPDLGAGGAPAPA-PVPAHTRDK-----DRQ 67

QY 73 GTANLAGDNGGNGGGRGGEGGRGNADATFTYRPSVAHRRARESPSLSDAIFKQSH 132
DB 68 TSV-----GDGHWELR-----CH-RLQDSLFSDDSGF-SNY 96
QY 133 AGLNLCVWVLIIVANSRLIENIMKYGLIRDFWFSRSLRD----WPLFMCCLSLSTIPP 189
DB 97 RGIILNWCVVMLILSNARLFLENLIKYGILV-DPIQVWSLFLKDPHWPAPCLIIASNFI 155
QY 190 LAATFVKEVLQKYLSEPVVFLHIIITMTVEVLVYVVTLCRDSAFSLGVTMLLT-CIV 248
DB 156 VATQIEKRISVGALTQOMGLLHVNLATIIICFPAAVALLVESITPVGSLFALASYSII 215
QY 249 WLKLVSYAHTSY-----DIRSLANAADK-----ANPEVSY--YVSLKSLAYFWAPTLC 295
DB 216 FLKFSYRDVNLWCRQRVAKAVSAGKVKVGAQAQNTVSPDNLYRDLVYFIFAPTL 275
QY 296 YQPSYPSACIRKGVARQAKLIVFTGFMGFIIEQYINPVRNSKHPKGLDLYA--IE 353
DB 276 YELNFRSPRIRKRELLRLLELFTQLQVGLIQQWVPTIQNSMKPKF-DMDYSRIIE 334
QY 354 RVLKLSVNPVWLCMFCFFHLWNLIAELLCFGDREFFYKDWNAKSGVDYWRMNMV 413
DB 335 RLLKLVNPVNLIIITMTVEVLVYVVTLCRDSAFSLGVTMLLT-CIV 473
QY 414 HKMVRHIYFPLRSKIPKTLAIIIAFLVSAVPHELCIAVPCRLFLKWLAFGLIMFOVPLV 473
DB 395 HKWCIRHFYKPEMLRGLSNKWMARTGVFLASAPHEVLSVPLRMFLMARTGMMAQIPLA 454
QY 474 FITNLOERGSTVGNMIFWTFICIFGQPMCVLLYYHD 511
DB 455 WIVNRF-----FQNGYNAAVW-VTLIIGOPVAVLVMVHD 488

RESULT 6
US-09-103-754A-5
; Sequence 5, Application US/09103754A
; Patent No. 6344548
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-acyltran
; TITLE OF INVENTION: sferase
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed
; STREET: 285 Hamilton Avenue, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/103,754A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Field, Bret E
; REGISTRATION NUMBER: 37,620
; REFERENCE/DOCKET NUMBER: 6510-105p
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650 327 3400
; TELEFAX: 650 327 3231
; TELEX:

Db 301 PRSACIRKGVARQAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIEVLKSLV 360
QY 361 PNLYVWLCMFYCFHFLWNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVRH 420
Db 361 PNLYVWLCMFYCFHFLWNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVRH 420
QY 421 IYFPCLRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKLWAFGLIMFQVPLVFTITNYLQ 480
Db 421 IYFPCLRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKLWAFGLIMFQVPLVFTITNYLQ 480
QY 481 ERFGSTVGNMIFWFIICIFGQPMCVLLYYHDLNMRKGSMS 520
Db 481 ERFGSTVGNMIFWFIICIFGQPMCVLLYYHDLNMRKGSMS 520

RESULT 2
US-09-593-359-4
; Sequence 4, Application US/09593359
; Patent No. 6552250
; GENERAL INFORMATION:
; APPLICANT: Laroché, Andre J.
; APPLICANT: Nykiforuk, Cory L.
; APPLICANT: Weselake, Randall J.
; TITLE OF INVENTION: Diacylglycerol O-acyltransferase
; FILE REFERENCE: 24015050
; CURRENT APPLICATION NUMBER: US/09/593,359
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 4
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Brassica napus
; OTHER INFORMATION: DGAT1
US-09-593-359-4

Query Match 83.3%; Score 2309.5; DB 4; Length 503;
Best Local Similarity 84.88; Pred. No. 8.2e-220;
Matches 442; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 1 MAILDSAGVTT-VTENGSGEFDLRLRRKRSRSDSSNGLLSGDNNSPSDVDGAPADY 59
Db 1 MAILDSGVAVPPTENG---VADLDRLHRRKSSDSSNGLL---SDTSPSDVGAAAE 53
QY 60 RDRDVSVDNDAGTANLAGDNGGDNNGGGGGGCGGNADATFYRPSVPAHRRARE 119
Db 54 RDRDVSAAEEAGTANLA-----GGDAETRESAGG-----DVRFTYRPSVPAHRRTRE 102
QY 120 SPLSSDAIFKQSHAGFLNLCVVLAVNSRLIENLMKYGLWLTDFWFSRSLRDWPLF 179
Db 103 SPLSSDAIFKQSHAGFLNLCVVLAVNSRLIENLMKYGLWLTDFWFSSTSLRDWPLF 162
QY 180 MCCISLSIFPLAFTVEKLVQKYESEPVVIFLHIITMTTEVLYPVVTVLRCDSAFLSGV 239
Db 163 MCCLSLSVFLAFTVEKLVQKYESEPVVAILHVIITMTTEVLYPVVTVLRCDSAFLSGV 222
QY 240 TMLLTCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPVTLCYQPS 299
Db 223 TMLLTCIVWLKLVSAHTSYDIRSLANSADKVDPEISYVSLKSLAYFMVAPVTLCYQPS 282
QY 300 YPSACIRKGVARQAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIEVLKLS 359
Db 283 YPSACIRKGVARQAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIEVLKLS 342
QY 360 VPVLYVWLCMFYCFHFLWNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVR 419
Db 343 VPVLYVWLCMFYCFHFLWNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVR 402
QY 420 HIYFPCRLRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKLWAFGLIMFQVPLVFTITNYL 479
Db 403 HVYFPCRLRIKIPKVPALIIIAFLVSAVHELCIAVPCRLFNLFMFMGIMFQVPLVFTITNEL 462
QY 480 QERFGSTVGNMIFWFIICIFGQPMCVLLYYHDLNMRKGSMS 520

Db 463 QERFGSMVGNMIFGSASCIFGQPMCGLLYYHDLNMRKGSMS 503

RESULT 3
US-09-593-359-2
; Sequence 2, Application US/09593359
; Patent No. 6552250
; GENERAL INFORMATION:
; APPLICANT: Laroché, Andre J.
; APPLICANT: Nykiforuk, Cory L.
; APPLICANT: Weselake, Randall J.
; TITLE OF INVENTION: Diacylglycerol O-acyltransferase
; FILE REFERENCE: 24015050
; CURRENT APPLICATION NUMBER: US/09/593,359
; CURRENT FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Brassica napus
; OTHER INFORMATION: DGAT2
US-09-593-359-2

Query Match 62.08; Score 1717; DB 4; Length 341;
Best Local Similarity 93.08; Pred. No. 1.8e-161;
Matches 317; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

QY 180 MCCISLSIFPLAFTVEKLVQKYESEPVVIFLHIITMTTEVLYPVVTVLRCDSAFLSGV 239
Db 1 MCCLSLSVFLAFTVEKLVQKYESEPVVIFLHIITMTTEVLYPVVTVLRCDSAFLSGV 60
QY 240 TMLLTCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPVTLCYQPS 299
Db 61 TMLLTCIVWLKLVSAHTSYDIRSLANSADKVDPEISYVSLKSLAYFMVAPVTLCYQPS 120
QY 300 YPSACIRKGVARQAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIEVLKLS 359
Db 121 YPSACIRKGVARQAKLVFTGFMGFIIEQYINPIVRNSKHPKLGDLLEYAIEVLKLS 180
QY 360 VPVLYVWLCMFYCFHFLWNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVR 419
Db 181 VPVLYVWLCMFYCFHFLWNLIAELLCFGDREFYKDWNNAKSVGDYWRWNNMPVHKWVR 240
QY 420 HIYFPCRLRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKLWAFGLIMFQVPLVFTITNYL 479
Db 241 HVYFPCRLRIKIPKVPALIIIAFLVSAVHELCIAVPCRLFNLFMFMGIMFQVPLVFTITNEL 300
QY 480 QERFGSTVGNMIFWFIICIFGQPMCVLLYYHDLNMRKGSMS 520
Db 301 QERFGSMVGNMIFGSASCIFGQPMCGLLYYHDLNMRKGSMS 341

RESULT 4
US-09-165-042-1
; Sequence 1, Application US/09165042
; Patent No. 6100077
; GENERAL INFORMATION:
; APPLICANT: Sturley, Stephen L.
; APPLICANT: Oelkers, Peter
; TITLE OF INVENTION: ISOLATION OF A GENE ENCODING DIACYLGLYCEROL
; FILE REFERENCE: 0575/56331
; CURRENT APPLICATION NUMBER: US/09/165,042
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 488
; TYPE: PRT
; ORGANISM: Yeast
US-09-165-042-1

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2003, 04:55:04 ; Search time 43 Seconds
(without alignments)
511.666 Million cell updates/sec

Title: US-09-623-514a-2

Perfect score: 2771

Sequence: 1 MAILDSAGVTTTENGGEF.....OPMCVLLYHDLNKRKGSMS 520

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

- 1: /cgn2.6/ptodata/1/1aa/5A_COMB.pep.*
- 2: /cgn2.6/ptodata/1/1aa/5B_COMB.pep.*
- 3: /cgn2.6/ptodata/1/1aa/6A_COMB.pep.*
- 4: /cgn2.6/ptodata/1/1aa/6B_COMB.pep.*
- 5: /cgn2.6/ptodata/1/1aa/PCTUS_COMB.pep.*
- 6: /cgn2.6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	520	4	US-09-326-203A-2
2	2309.5	83.3	503	4	US-09-593-359-4
3	1717	62.0	341	4	US-09-593-359-2
4	791.5	28.6	488	3	US-09-165-042-1
5	773.5	27.9	500	4	US-09-326-203A-17
6	770.5	27.8	498	4	US-09-103-754A-5
7	724.5	26.1	496	4	US-09-326-203A-18
8	696.5	25.1	386	4	US-09-103-754A-4
9	413	14.9	409	4	US-09-326-203A-23
10	396	14.3	550	1	US-08-121-057-4
11	396	14.3	550	2	US-08-509-187D-4
12	396	14.3	550	2	US-09-121-396-4
13	396	14.3	550	5	PCT-US93-09704A-4
14	375	13.5	525	4	US-09-328-857A-3
15	361	13.0	432	4	US-09-326-203A-25
16	357.5	12.9	522	3	US-09-165-042-3
17	351	12.7	502	4	US-09-328-857A-4
18	298.5	10.8	429	3	US-09-326-203A-24
19	190	6.9	51	3	US-09-165-042-9
20	153.5	5.5	52	3	US-09-165-042-6
21	147	5.3	53	3	US-09-165-042-8
22	143.5	5.2	412	4	US-09-134-001C-4885
23	143	5.2	53	3	US-09-165-042-7
24	141	5.1	52	3	US-09-165-042-5
25	123	4.4	250	4	US-09-326-203A-22
26	122.5	4.4	455	4	US-09-198-452A-1028
27	120	4.3	1235	1	US-08-118-101A-2

28	119	4.3	30	3	US-09-165-042-17	Sequence 17, Appl
29	112	4.0	352	4	US-09-205-815B-42	Sequence 42, Appl
30	102.5	3.7	33	3	US-09-165-042-10	Sequence 10, Appl
31	102.5	3.7	289	3	US-09-184-658-63	Sequence 63, Appl
32	101	3.6	813	3	US-08-836-325-8	Sequence 8, Appl
33	100	3.6	333	4	US-09-170-496D-8	Sequence 8, Appl
34	100	3.6	333	4	US-09-170-496D-168	Sequence 168, App
35	98	3.5	320	4	US-09-439-313-527	Sequence 527, App
36	97.5	3.5	461	2	US-08-463-587A-26	Sequence 26, Appl
37	97.5	3.5	461	3	US-08-463-667A-4	Sequence 4, Appl
38	97.5	3.5	461	3	US-08-923-854-26	Sequence 26, Appl
39	97.5	3.5	461	5	PCT-US91-09133-27	Sequence 27, Appl
40	97.5	3.5	975	4	US-09-328-352-4764	Sequence 4764, Ap
41	96	3.5	2237	1	US-08-455-543A-48	Sequence 48, Appl
42	96	3.5	2237	2	US-08-223-305C-48	Sequence 48, Appl
43	96	3.5	2339	1	US-08-455-543A-47	Sequence 47, Appl
44	96	3.5	2339	2	US-08-223-305C-47	Sequence 47, Appl
45	95.5	3.4	384	3	US-09-071-434-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-09-326-203A-2

; Sequence 2, Application US/09326203A

; Patent No. 6444876

; GENERAL INFORMATION:

; APPLICANT: Lassar, Mike

; APPLICANT: Ruzinsky, Diane

; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic

; FILE REFERENCE: 17045/00/WO

; CURRENT APPLICATION NUMBER: US/09/326, 203A

; CURRENT FILING DATE: 1999-06-04

; PRIOR APPLICATION NUMBER: 60/088,143

; PRIOR FILING DATE: 1998-06-05

; PRIOR APPLICATION NUMBER: 60/108,389

; PRIOR FILING DATE: 1998-11-12

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 520

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-326-203A-2

Query Match 100.0%; Score 2771; DB 4; Length 520;
Best Local Similarity 100.0%; Pred. No. 2e-265;
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MAILDSAGVTTTENGGEFVLDRLRRKRSRSDSSNGLLSGSDNNSPSDDVGAPADVR	60
DB	1	MAILDSAGVTTTENGGEFVLDRLRRKRSRSDSSNGLLSGSDNNSPSDDVGAPADVR	60
QY	61	DRIDSVDNDAQGTANLAGDNNGGGGRGGGGRGNADATFTYRSPVAHRRARES	120
DB	61	DRIDSVDNDAQGTANLAGDNNGGGGRGGGGRGNADATFTYRSPVAHRRARES	120
QY	121	PLSSDAIFKQSHAGLFLNLCVVLIIVNSRLIENLMKYGLWLTDFWFSRSLRDPLEPM	180
DB	121	PLSSDAIFKQSHAGLFLNLCVVLIIVNSRLIENLMKYGLWLTDFWFSRSLRDPLEPM	180
QY	181	CCISLSIFPLAAFTVEKLVQKYISEPVVIFLHIITMTVEVLPVVTURCDSAFLSGVT	240
DB	181	CCISLSIFPLAAFTVEKLVQKYISEPVVIFLHIITMTVEVLPVVTURCDSAFLSGVT	240
QY	241	LMLLTCIVWLKLVYAHTSYDIRSLANAADKANPEVSYVYVSLKSLAYFMVAPTLQYPSY	300
DB	241	LMLLTCIVWLKLVYAHTSYDIRSLANAADKANPEVSYVYVSLKSLAYFMVAPTLQYPSY	300
QY	301	PRSACIRKQWVARQAKLVIFTFGFWGFIIEQVFNPIVRNSKPKLGLDLYATERYVLKLSV	360
DB	301	PRSACIRKQWVARQAKLVIFTFGFWGFIIEQVFNPIVRNSKPKLGLDLYATERYVLKLSV	360


```
RESULT 12
US-10-157-855-18
; Sequence 18, Application US/10157855
; Publication No. US20020170091A1
; GENERAL INFORMATION:
; APPLICANT: Lassner, Michael W.
; APPLICANT: Ruzinsky, Diane M.
; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic
; FILE REFERENCE: Acid Sequences
; CURRENT APPLICATION NUMBER: US/10/157,855
; PRIOR FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 09/326,203
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,143
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/108,389
; PRIOR FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 18
; LENGTH: 496
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-157-855-18

Query Match      26.1%; Score 724.5; DB 14; Length 496;
Best Local Similarity 31.5%; Pred. No. 2.4e-59;
Matches 166; Conservative 98; Mismatches 166; Indels 97; Gaps 14;

QY 27 RRRKRSRSSGGLLSGSDNNSPSDDVGDAPADVRDRIDSVVNDDAQGTANLAGDNNGGD 86
Db 7 RRRQPSSTSGSLASSR-----RSSEAQ 30

QY 87 NNGGGRGGEGGNADFTYRPSVPAHRRARESPSSDALSFKQSHA-----GLENLCVW 141
Db 31 NNSRSKSSRMGCKEV-----VH-----TAQDSLSTSGGWNFRGFNLSIL 75

QY 142 VLIAVNSRLIENLMKYGLIRTDWFSS-----RSLRDWLFMCICISIFPLAAFTVE 196
Db 76 LVLNSGRVALENVIKYLITPLQWISTFEVHHYSIWSNPNALILCSNQLISVFGME 135

QY 197 KLVQKYLSEPVIFLHIITMTEVLYPVYVTLRCD-SAFISGVTLMLLTCTIVWLKLVSY 255
Db 136 KILRGWLGNGFAAVFTSLVIAHLTIPVYVTLTHKKNPLWSVVMGVVIEALKFTSY 195

QY 256 AHTSY---DIR-----SLANADKA-----NPEVSY-----VSLKSLA 286
Db 196 GHVNTWARDARKTELKTQVTDLAKTCDPKQFMDLKDLSMHQMAAQAQYPAINTLSNIY 255

QY 287 YFMVAPTLCYQPSYPRSAIRKGVARQAKLIVFTGPMGTIEQYINPIVRNSKHPLKG 346
Db 256 YFMAAPTLCYEFKPRRLRIRKHLIKRTVELIFLSFLIALVQVQVVTYVNSMKPL-S 314

QY 347 DLLYA--IERVKLSVPNLYWLCMFYCFHLMNLIAELLCFGDREYFKDWNNAKSVGD 404
Db 315 EMEYSRCLERLLKLAIPNHLIWLFFYTFHFSFLNLAELLRFADREFYRDFWNAETIGY 374

QY 405 YWRMNMNPKHMYRHHVFPCLRSKIPKTLAIITAFVLSAVFHELCIAVPCRLKLAFL 464
Db 375 FWKGNINPVRFAVHHVYRMMNNSKMSAFFVVFVFAFFHELYSVPLKIPFLWSY 434

QY 465 GIMEQVPLVFTNTVQERFGSTGVNMIWFIFCFGPMCVLLYYHD 511
Db 435 GMMGQIPLSIITDKVR--GGRTGNIIVW--LSLIVGQPLAILMYGHD 478

RESULT 13
US-10-273-438-5
; Sequence 5, Application US/10273438
; Publication No. US20030072757A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V.
```

```
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-Acyltransferase
; FILE REFERENCE: UCAL-105CIP2
; CURRENT APPLICATION NUMBER: US/10/273,438
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US/10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 386
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-273-438-5

Query Match      25.1%; Score 696.5; DB 15; Length 386;
Best Local Similarity 40.9%; Pred. No. 7.4e-57;
Matches 157; Conservative 68; Mismatches 126; Indels 33; Gaps 12;

QY 150 LIENLMKYGLIRTDWFSSSLRD---WPLFMCICISIFPLAAFTVEKLVLYQYISE 206
Db 1 LLENLIRYGLIV-DPIQVSLFLKDPYSWAPCLVIAANVPAVAFAVQVEKRLAVGALTE 59

QY 207 PVVIFLHIITMTEVLYPVYVTLRCDSAFLSGVTLMLLT-CIVWLKLVSYAHTSYDIRS- 264
Db 60 QAGLLHLVANLTIICFPAAVVLVESITPVGSLALMAHTILFLKLSYR---DVNSW 115

QY 265 -----LANADKANP-----EVSY--YVLSKSLAYFMVAPTLCYQPSYPRSAIRKGV 310
Db 116 CRRARAKAASAGKASSVAAPHTVSYPDNLTYRDLVYFLFAPTLCYELNFPSPRIKRF 175

QY 311 VARQAKLIVFTGPMGTIEQYINPIVRNSKHPLKGLLYA--IERVKLSVPNLYWLC 368
Db 176 LLRLTLEMLFTQLQVGLIQQWVPTIQQNSMKPKF-DMDYSRIERLLKLAVPNHLIWL 234

QY 369 MEYCFEHLNLTIAELLCFGDREYFKDWNNAKSVGDYWRMNMNPKHMYRHHVFPCLRS 428
Db 235 FFYWLFSCLNAVAELMQDREFYRDMWNSSEVTYFWQNNIPVHKWCIRHFYKPMCLRR 294

QY 429 KIPKTLAIITAFVLSAVFHELCIAVPCRLKLAFLGIMFQVPLV-FITNYLQERFGSTV 487
Db 295 GSSKWMATGTVFLASAFHELYSVPLRMFLMAFTGMAQIPLAFLWYGRFGQNY---- 350

QY 488 GNMIFWTFICIFGPMCVLLYYHD 511
Db 351 GNAAVW--LSLIIGQPIAVLMYVHD 373

RESULT 14
US-09-764-853-499
; Sequence 499, Application US/09764853
; Patent No. US20020090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PU206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 499
; LENGTH: 236
```

```

RESULT 11
US-10-273
; Sequenc
; Publica
; GENERAL
; APPLIC
; APPLIC
; APPLIC
; APPLIC
; TITLE
; FILE R
; CURREN
; CURREN
; PRIOR
; PRIOR
; PRIOR
; PRIOR
; PRIOR
; PRIOR
; PRIOR
; PRIOR
; NUMBER
; SOFTWA
; SEQ ID
; LENGTH
; TYPE:
; ORGAN
US-10-273

```

[illegible]

```

RESULT 10
US-10-273-438-7
; Sequence 7, Application US/10273438
; Publication No. US20030072757A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V.
; APPLICANT: Cases, Sylvaine
; APPLICANT: Smith, Steven
; APPLICANT: Erickson, Sandra
; TITLE OF INVENTION: Diacylglycerol O-Acyltransferase
; FILE REFERENCE: UCAL-105CIP2
; CURRENT APPLICATION NUMBER: US/10/273,438
; CURRENT FILING DATE: 2002-10-16
; PRIOR APPLICATION NUMBER: US/10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 498
; TYPE: PRT
; ORGANISM: mus musculus
US-10-273-438-7

```

Query Match	27.88;	Score	770.5;	DB	15;	Length	498;
Best Local Similarity	36.3%;	Pred. No.	1.2e-63;				
Matches	190;	Conservative	83;	Mismatches	179;	Indels	71;
	Gaps	15;					
QY	8	GVTVTENGGEFVDLRLRRKRSRSDSSNGLLLSGSDNNSPSDVGCAPADVRDRIDSV	67				
Db	16	GSRSVGGSGPKVEEDVRDAVSPD-----LCAGGDAPAP-----APAUTRK-----	62				
QY	68	NDDAQGTANLAGDNGGDDNGGGGCGEGRGGNADATFTYRPSYPAHRRARESPSSDAI	127				
Db	63	-----DGRTSVGDGWDLR-----HRLQDSLFSDDSG	90				
QY	128	FKOSHAGLNLVCVVLIATNSRLIENLMKYGWLIRTDWFESSRLRD---WPLFMCCIS	184				
Db	91	F-SNYRGILNCWVMLLSNARLFENLIRKIGILV-DPIQVSLFLKDPYSWPAPCVIIA	148				
QY	185	LSIFPLAAAFVEKVLQKYISEPVWIFLHIITMTEVLVYVYVTLRCDSAFLSGVTLMLL	244				
Db	149	SNLFVNAQFIERLAVGAUTEQWGLLHVVNLTATICFPAAVALLVESTPTVGSVFLA	208				
QY	245	T-CIVMLKLYSAHTSY-----DIRSLANAADK-----ANPEYSY--YVSLKSLAYFMV	290				
Db	209	SYSIMFLKLYSRDVLNLCWRQRRVKAKAVSTGKKVSGAAQAQVSYPDNLTYRDIYFIF	268				
QY	291	APLTCQPSYPSRACIRKGWAKQFAKLVTFTGMFGFIIBQYINPIVNSKHPILKGDLLY	350				
Db	269	APTLCYELNPPSRPIRKRELLRRVLEMLFPTQLQVGLIQWVPTIHSNKKPEK-DMDY	327				
QY	351	A--IERVLKLSVNLVYVLMCFEYFPHLNIILAEELCFGDREYFKDWNNAKSVGDDYRM	408				
Db	328	SRRIERLLKUANPHLLIWLITFFWFFHSCUNAVAEELQFGDREYRDWNNAESYTFWQN	387				
QY	409	WNMPVHKWVRHYTFPCRLSKIPKTLTAIIATFLVSAVFHELCIAVPCRLKFLWAFGLIME	468				
Db	388	WNIPVHKWCIRHEYKPMLRHSGSKWARTGVLTSAFFHEYVLVSVPLRMFLWAFATAMA	447				
QY	469	QVPLVITNYLQERFSGTVGNMIFWIFFCIFGQPMCVLLYYHD	511				
Db	448	QVPLAWTVGRF----FOGNYCNAAV--VTLIIIGOPAVLMVYHD	486				

[illegible]

```

RESULT 6
US-10-223-076-13
; Sequence 13, Application US/10223076
; Publication No. US20030074695A1
; GENERAL INFORMATION:
; APPLICANT: Farese, Robert V
; APPLICANT: Cases, Sylvaine
; TITLE OF INVENTION: Plant diacylglycerol O-transferase and
; TITLE OF INVENTION: Uses thereof
; FILE REFERENCE: UCAL-105CIP3
; CURRENT APPLICATION NUMBER: US/10/223,076
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Nicotiana tabacum
US-10-223-076-13

```

	Query Match	66.5%; Score 1842.5; DB 15; Length 532;
	Best Local Similarity	67.0%; Pred. No. 5.4e-164;
	Matches	358; Conservative 49; Mismatches 104; Indels 23; Gaps 7;
QY	1 MAILDSAGVTTVTENGGEFVDL---	RLRRRKRSRSDSSGLLLSGDNNSPDSDVGAPA 57
DB	4 MELPSEVMETTTTTTSGIENLNLDLHNSVR---	GSGFEAASAINSS---DANMSE 56
QY	58 DYVRDRIDS-----VVNDDAQGTANLADNGNDDNGGGEGCGCNAUAT---	PXYR 108
DB	57 DRDVDCSGANGLFTVNERSKSVGESDVIRKEDRDNDNVANGESKSTETTTPFFKPAYR	116
QY	109 PSVPAHRRESPLSSDAIFKQSHAGLFNLGVVLIAVNSRLLIENLMKYGLIRTDFWF	168
DB	117 ASAPAHRRIKESPLSSDAIFKQSHAGLFNLGVVLIAVNSRLLIENLMKYGLIRAGFWF	176
QY	169 SRSRLRDWPLEMCISLSIFFLAAFTVEKLVLQAKISEPVVIFHLIIITTEVLVPYYVT	228

```

177 SSKSLRDWPLLMCCLSQLPLAFLVEKLAQORHLTERAVVTLHTITTTAAILYPVLVI 233
229 LRCDSAFSLGVTMLLTCTIYWLKLVSAHTSYDILSLANAADK---ANPEYSYYVSLKSL 285
237 LGCDSAFELFGVILMLVACIYWMKLVSAHTNHRQLAKSTYDKDTSYDVSFVSKSL 296
286 AYFMVAPTLCYQPSYPSRACIRKGWVARQFAKLVIFTGFMGFIIEQYINPIVNRNSKHPLK 345
297 AYFMVAPTLCYQLSPYHTPCIRKKGWVARQFIKLVIFTGLMGFIIEQYINPIVONSOHPLK 356
346 GDLLYATERYVLKLSVPNLVYWLCMFYCFHFHMLNLIAELLCFGDRFEYKDMWNAKSGVDY 405
357 GNLLYATERYVLKLSVPNLVYWLCMFYCFHFHMLNLIAELLCFGDRFEYKDMWNAKTIDEBY 416
406 WRWNNMPVHKWVWPHIYFPCILRSKIPKTLAIILIAFLVSAVPHELCIAVPCRLFKLWAFILG 465
417 WRWNNMPVHKWVWPHIYFPCILRNGIPKGVAILIIAFLVSAVPHELCIAVPCRLFKWAFMG 476
466 IMFQVPLVFITNYLQERF-GSTGVNMIFWFICFGQPMCVLLYYLHDLMMNRKGS 518
477 IMFQVPLVLTNLFNLQNFQSSVGNMFMWCFPCILGQPMCVLLYYHDMNRKSS 530

RESULT 7
US-10-223-076-7
; Sequence 7, Application US/10223076
; Publication No. US20030074695A1
; GENERAL INFORMATION:
; APPLICANT: Faresse, Robert V
; APPLICANT: Cases, Sylvaine
; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: UCAL-105CIP3
; CURRENT APPLICATION NUMBER: US/10/223,076
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Brassica napus
US-10-223-076-7

```

Query Match	62.04;	Score 1717;	DB 15;	Length 341;
Best Local Similarity	93.09;	Pred. No.1.7e-152;		
Matches 317; Conservative	10;	Mismatches 14;	Indels 0;	Gaps 0
QY	180	MCCLSISIFPLAAFTVEKVLQKIVISEPVVIFLHIIITMTTEVLYPVVYVTLRCDSAFLSGV	239	
Db	1	MCCLSISIFPLAAFTVEKVLQKIVISEPVVIFLHIIITMTTEVLYPVVYVTLRCDSAFLSGD	60	
QY	240	TLMLLTCTIVMLKIVSYAHTSYDIIISLANAAKANPEVSYVSLKSLAYFMVAPTLCYCPS	299	
Db	61	TLMLLTCTIVMLKIVSYAHTNYDIITLANSDDKANPEVSYVSLKSLAYFMVAPTLCYCPS	120	
QY	300	YPSACIRKKGWVAQFAKLVFTGFMGFIIEQYINPIVRNSKHPLKGLDLLYAIERVKLKLS	359	
Db	121	YPSPCIRKKGWVAQFAKLVFTGLMGFIIEQYINPIVRNSKHPLKGLDLLYAIERVKLKLS	180	
QY	360	VPNLYVMVCMFYCFPHLWNLNTLAEILCFDREFYKDWNNAKSVGDYWRMNNPVPVKWMYR	419	
Db	181	VPNLYVMVCMFYCFPHLWNLNTLAEILCFDREFYKDWNNAKSVGDYWRMNNPVPVKWMYR	240	

121	Qy	PLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIENLMKYGLWLTIRDTWFSSRSLRDWPLFM	180
121	Db	PLSSDAIFKQSHAGLFNLCVVVLIAVNSRLIENLMKYGLWLTIRDTWFSSRSLRDWPLFM	180
181	Qy	CCISLSIFPLAAAFVTEKLVLOKYLISEPVVIFLHIIITMTVELVYPVVTILRCDSAFLSGVT	240
181	Db	CCISLSIFPLAAAFVTEKLVLOKYLISEPVVIFLHIIITMTVELVYPVVTILRCDSAFLSGVT	240
241	Qy	LMLLTCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVLSKSLAYFMVAFTLCYQPSY	300
241	Db	LMLLTCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVLSKSLAYFMVAFTLCYQPSY	300
301	Qy	PRSAICIRKGWVARQAKLVIPTGFMGFTIEQYINPIVRNSKHLPGDLLYATIERVLKLSV	360
301	Db	PRSAICIRKGWVARQAKLVIPTGFMGFTIEQYINPIVRNSKHLPGDLLYATIERVLKLSV	360
361	Qy	PNLYVWLCMFCFFPHLNLINLAELLCCGDRFEYKDWNNAKSVGDDYWRMNMNVHKKWVRH	420
361	Db	PNLYVWLCMFCFFPHLNLINLAELLCCGDRFEYKDWNNAKSVGDDYWRMNMNVHKKWVRH	420
421	Qy	IYFPCLRASKTPKTLAIITAFVLVSVAFFHELCTAVPCRLEFKLNAFLGIMFQVPLVFTNTYIQ	480
421	Db	IYFPCLRASKTPKTLAIITAFVLVSVAFFHELCTAVPCRLEFKLNAFLGIMFQVPLVFTNTYIQ	480
481	Qy	ERFGSTVGNMIFWFIFCIFGQPMCVLLYHYHDLNMRKSGMS	520
481	Db	ERFGSTVGNMIFWFIFCIFGQPMCVLLYHYHDLNMRKSGMS	520

```

RESULT 2
US-10-223-076-3
; Sequence 3, Application US/10223076
; Publication No. US20030074695A1
; GENERAL INFORMATION:
; APPLICANT: Faresse, Robert V
; APPLICANT: Cases, Sylvainne
; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: UCAL-105CIP3
; CURRENT APPLICATION NUMBER: US/10/223,076
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-223-076-3

```

Db	221	PLSSDAIFKOSHAGFLNLCVVVLVAVNSRLIENLMKYGLWIRITDFWFSRSLRDWPFLM	180
Qy	181	CCISLSIFPLAAFTVEKLVLOKYISEPVIFLHIIITWTEVLVYVYVTLRCDSAFLSGVT	240
Db	181	CCISLSIFPLAAFTVEKLVLOKYISEPVIFLHIIITWTEVLVYVYVTLRCDSAFLSGVT	240
Qy	241	LMLTLCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPTLCYQPSY	300
Db	241	LMLTLCIVWLKLVSAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPTLCYQPSY	300
Qy	301	PRSAICIRKGWVARQFAKLVIFTGPMGFIIIEQYINPIVRNSKHPKLGDLLYAIERVLKUSV	360
Db	301	PRSAICIRKGWVARQFAKLVIFTGPMGFIIIEQYINPIVRNSKHPKLGDLLYAIERVLKUSV	360
Qy	361	PNLYVWLJCMFCFPHLWNLITLAEULCFCGDRFYYKDWNAKSVGDYWRMNNPVRHKWVYRH	420
Db	361	PNLYVWLJCMFCFPHLWNLITLAEULCFCGDRFYYKDWNAKSVGDYWRMNNPVRHKWVYRH	420
Qy	421	IYFPCLRKSKIPKTLAIIAFLVSVAFFHELCLTAVPCRLFKLWAFLGIMFQVPLVFTINYLQ	480
Db	421	IYFPCLRKSKIPKTLAIIAFLVSVAFFHELCLTAVPCRLFKLWAFLGIMFQVPLVFTINYLQ	480
Qy	481	ERGSTVGNMIFWPIFCIFGQPMCVLLYYHDLMLNRKGSMS	520
Db	481	ERGSTVGNMIFWPIFCIFGQPMCVLLYYHDLMLNRKGSMS	520

```

RESULT 3
US-10-223-076-9
; Sequence 9, Application US/10223076
; Publication No. US20030074695A1
; GENERAL INFORMATION:
; APPLICANT: Farsee, Robert V
; APPLICANT: Cases, Sylvaine
; TITLE OF INVENTION: Plant Diacylglycerol O-transferase and
; FILE OF INVENTION: Uses Thereof
; FILE REFERENCE: UGAL-105CIP3
; CURRENT APPLICATION NUMBER: US/10/223,076
; CURRENT FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 10/040,315
; PRIOR FILING DATE: 2001-10-29
; PRIOR APPLICATION NUMBER: 09/339,472
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/107,771
; PRIOR FILING DATE: 1998-11-09
; PRIOR APPLICATION NUMBER: PCT/US98/17883
; PRIOR FILING DATE: 1998-08-28
; PRIOR APPLICATION NUMBER: 09/103,754
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Brassica napus
US-10-223-076-9

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2003, 04:59:04 ; Search time 61 seconds
(without alignments)
1166.009 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771

Sequence: 1 MAILDSAGVTTVTENGGEF.....QPMCVLLYHDLNRKGSMS 520

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 510680 seqs, 13678180 residues

Total number of hits satisfying chosen parameters: 510680

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published_Applications_AA:*
- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
 - 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
 - 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
 - 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
 - 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
 - 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
 - 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
 - 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
 - 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
 - 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
 - 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
 - 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
 - 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
 - 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
 - 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
 - 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
 - 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
 - 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	520	14	US-10-157-855-2
2	2771	100.0	520	15	US-10-223-076-3
3	2309.5	83.3	503	15	US-10-223-076-9
4	1880	67.8	518	15	US-10-223-076-11
5	1854	66.9	534	15	US-10-223-076-15
6	1842.5	66.5	532	15	US-10-223-076-13
7	1717	62.0	341	15	US-10-223-076-7
8	791.5	28.6	488	15	US-10-273-438-6
9	773.5	27.9	500	14	US-10-157-855-17
10	770.5	27.8	498	15	US-10-273-438-7
11	770.5	27.8	498	15	US-10-273-438-10
12	724.5	26.1	496	14	US-10-157-855-18
13	696.5	25.1	386	15	US-10-273-438-5
14	573.5	20.7	236	9	US-09-764-853-499
15	413	14.9	409	14	US-10-157-855-23

16	397	14.3	550	12	US-10-303-664A-12	Sequence 12, Appl
17	394.5	14.2	409	14	US-10-157-855-22	Sequence 22, Appl
18	361	13.0	432	14	US-10-157-855-25	Sequence 25, Appl
19	340.5	12.3	165	10	US-09-764-868-1200	Sequence 1200, Ap
20	298.5	10.8	429	14	US-10-157-855-24	Sequence 24, Appl
21	257	9.3	74	11	US-09-969-730-155	Sequence 155, Appl
22	257	9.3	75	11	US-09-774-639-125	Sequence 125, App
23	214	7.7	163	11	US-09-774-639-230	Sequence 230, App
24	214	7.7	163	11	US-09-969-730-298	Sequence 298, App
25	188.5	6.8	109	9	US-09-764-853-734	Sequence 734, App
26	187.5	6.8	60	11	US-09-774-639-237	Sequence 237, App
27	187.5	6.8	60	11	US-09-969-730-305	Sequence 305, App
28	121	4.4	48	11	US-09-774-639-238	Sequence 238, App
29	121	4.4	48	11	US-09-969-730-306	Sequence 306, App
30	112	4.0	579	10	US-09-922-364A-19	Sequence 19, Appl
31	112	4.0	579	10	US-09-254-590-19	Sequence 19, Appl
32	112	4.0	579	14	US-10-115-695-19	Sequence 19, Appl
33	112	4.0	579	15	US-10-116-561-19	Sequence 19, Appl
34	112	4.0	579	15	US-10-115-671-19	Sequence 19, Appl
35	112	4.0	579	15	US-10-115-415-19	Sequence 19, Appl
36	112	4.0	579	15	US-10-116-260-19	Sequence 19, Appl
37	112	4.0	579	15	US-10-115-688-19	Sequence 19, Appl
38	112	4.0	847	11	US-09-875-321-2	Sequence 2, Appl
39	112	4.0	847	15	US-10-162-012-2	Sequence 2, Appl
40	111	4.0	420	15	US-10-127-032-92	Sequence 92, Appl
41	110.5	4.0	433	15	US-10-204-887-100	Sequence 100, App
42	109.5	4.0	388	11	US-09-964-923A-15	Sequence 15, Appl
43	109.5	4.0	388	15	US-10-225-567A-318	Sequence 318, App
44	108	3.9	527	10	US-09-881-752A-136	Sequence 136, App
45	106	3.8	506	9	US-09-804-551B-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1
US-10-157-855-2
; Sequence 2, Application US/10157855
; Publication No. US20020170091A1
; GENERAL INFORMATION:
; APPLICANT: Lasser, Michael W.
; APPLICANT: Ruzinsky, Diane M.
; TITLE OF INVENTION: Acyl-CoA:Cholesterol Acyltransferase Related Nucleic
; FILE OF INVENTION: Acid Sequences
; FILE REFERENCE: 16516.158
; CURRENT APPLICATION NUMBER: US/10/157,855
; CURRENT FILING DATE: 2002-05-31
; PRIOR APPLICATION NUMBER: 09/326,203
; PRIOR FILING DATE: 1999-06-04
; PRIOR APPLICATION NUMBER: 60/088,143
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/108,389
; PRIOR FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 520
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-10-157-855-2

Query Match	100.0%	Score 2771;	DB 14;	Length 520;
Best Local Similarity	100.0%;	Pred. No. 5.8e-251;		
Matches 520;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	MAILDSAGVTTVTENGGEFVLDLRLRRKRSRSDSSNGLLSGSDNNSPSDVGGAPADV	60	
DB	1	MAILDSAGVTTVTENGGEFVLDLRLRRKRSRSDSSNGLLSGSDNNSPSDVGGAPADV	60	
QY	61	DRIDSVVNDDAQGTANTLAGDNNGGDNNGGGCGGEGRGGNADATFTYRPSVPAHRRARES	120	
DB	61	DRIDSVVNDDAQGTANTLAGDNNGGDNNGGGCGGEGRGGNADATFTYRPSVPAHRRARES	120	

THIS PAGE BLANK (USPTO)

QY 298 PSYPSACIRKGVARQFAKLVITGFCMGFIIEOYINPIVRNSKHPKGLDLYA--IERV 355
 Db 180 LNPSPRIRKFLRLRILEMFFQQLQVLQOQWMPVPTIQNSMKPFK-DMDSYRIERL 238
 QY 356 LKLSVPLNVLWCMFCFPHLWNLIALLCFGDREYFKDWNNAKSGVDYWRWNNPVHK 415
 Db 239 LKLVNPHLILWIFFYFWLPHSCNVAELMQGDFREYFDWNWSESVTFWQWNNIPVK 298
 QY 416 WVRHYHYPCLSKIPKTKLIIAIIIVLSVAFVHELCAVPCRLFKLWAFGLIMFQVPLV-F 474
 Db 299 WCIRHFYKPMRLRGSSKMMARTGVFLASAFFEYLVSVPLRMFLWAFMGMAQIPLAWF 358
 QY 475 IYNYLQERGSGVGNWIFWIFCIGOPCMVLLIYHD 511
 Db 359 VGRFTQGN-----GNAAVW-LSLIIGQPIAVLWYHD 390
 RESULT 14
 Q96008
 ID Q96008 PRELIMINARY; PRT: 565 AA.
 AC Q96008;
 DT 01-DEC-2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)
 DE LD33852P (CG31991-PA)
 GN MDY OR CG13273 OR CG17937 OR CG17938 OR CG31991.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuncio J., Pacleib J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Xu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]

RA Svirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.X., Wasserman D.A., Weinstein K.C., Wu D., Weissbach J.,
 RA Williams S.M., Woodager, Worley K.C., W. D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferrizera S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuncio J.,
 RA Pacleib J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirkas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Klomp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY051835; RAK93259.1; -
 DR EMBL; AF003652; RAN10972.1; -
 DR FlyBase; FBgn0004797; mdv.
 DR InterPro; IPR001064; Crystallin.
 DR InterPro; IPR004299; MBOAT_fam.
 DR Pfam; PF03062; MBOAT; 1.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
 SQ SEQUENCE 565 AA; 64943 MW; 118B2AF6A385C569 CRC64;
 Query Match 26.2%; Score 725.5; DB 5; Length 565;
 Best Local Similarity 32.4%; Pred. No. 1.6e-52;
 Matches 179; Conservative 96; Mismatches 202; Indels 75; Gaps 14;
 QY 13 TENGGEFFVD-LDLRRKRSRSDSNGLLSGSDNNPSDDYVGAADVRDRSDSVNDDA 71
 Db 19 TANSSSGVGIMKLRSSASATEHNLSSLRNKSTQNLFDHGNPIDLR-QYRKVLDK-- 75
 QY 72 OCTANLADGNNGGDNGGGRGGGGR-----GNADATFTYRPSVPAHR 115
 Db 76 -----DENGNGTNGSEKKLRYRTQSVTRAETSNEKEKQRAQPGRIH- 120
 QY 116 RARESPLSDAIFKQSHAGLNLNLCVVLVIANVRLIENLKYGLWIRTFDFWSSRLRD 175
 Db 121 RPRLSLFWSWGGF-TNFSGLVNWGFLLLCIGGLLENLKYGIRINPLDFWFFISGHN 179
 QY 176 WPLEMCCISLSIFPLA----AFTVEKLVLQKYSVPVWIFLHITMTFVLPVYVTLRC 231
 Db 180 EGEHGNALISLISLVHLSLCLAVEKGLAMEIIAEGGLGFLQIVNIVLVLCPLV-VTIIHL 238

	Query Match	26.7%	Score 739.5	DB 6	Length 489
Best Local Similarity	35.8%				
Matches 182	Conservative 82				
	Mismatches 181				
	Indels 63				
	Gaps 15				

[illegible]

RA Nylkiforuk C.L., Laroche A., Weselake R.J.;
 RT "Isolation and Sequence Analysis of a Novel cDNA Encoding a Putative
 RT Diacylglycerol Acyltransferase from a Microspora-derived Cell
 RT Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No.
 RT AF155224). (PGR99-123).";
 RL Plant Physiol. 120:1207-1207(1999).
 DR EMBL; AF155224; AAD40861.1; -;
 KW Acyltransferase; Transferase.
 SQ SEQUENCE 341 AA; 39532 MW; 1A46340C49F16332 CRC64;

Query Match 62.0%; Score 1717; DB 10; Length 341;
 Best Local Similarity 93.0%; Pred. No. 3.7e-136;
 Matches 317; Conservative 10; Mismatches 14; Indels 0; Gaps 0;

Qy 180 MCCISLSIFPLAFTVEKLVKQKISPEVVFILHIITMTVEVLYPVVYVTLRCDSAFLSGV 239
 Db 1 MCCISLSIFPLAFTVEKLVKQKISPEVVFILHIITMTVEVLYPVVYVTLRCDSAFLSGD 60

Qy 240 TMLLTCTVWLKLVSYAHTSYDRLSRLANADKANPEVSYVSLKSLAYFWVAPLTCYOPS 299
 Db 61 TMLLTCTVWLKLVSYAHTSYDRLSRLANADKANPEVSYVSLKSLAYFWVAPLTCYOPS 120

Qy 300 YPSACIRKGVARQFAKLVFTGFMGFIIEQYINPIVRNSKHPKGLDLYAIERVLKLS 359
 Db 121 YPSACIRKGVARQFAKLVFTGFMGFIIEQYINPIVRNSKHPKGLDLYAIERVLKLS 180

Qy 360 VPVLYVWLCMFYCFHFLWNLIAELLCFGDRFEYKDWNAKSVGDYWRMNMVPHKVMVR 419
 Db 181 VPVLYVWLCMFYCFHFLWNLIAELLCFGDRFEYKDWNAKSVGDYWRMNMVPHKVMVR 240

Qy 420 HIYPCLRSKIPKTLAIIIAFLYSAVFHELCIAVPCRLFKLWAFGLMFOVPLVFIINYL 479
 Db 241 HIYPCLRSKIPKTLAIIIAFLYSAVFHELCIAVPCRLFKLWAFGLMFOVPLVFIINFL 300

Qy 480 QERFGSTVGNMIFWFIICFPGCMVLLVYHDLNRRKGSMS 520
 Db 301 QERFGSVGNMIFGASCFIQPMCGGLYHDLNRRKGSMS 341

RESULT 8
 Q96BB8 PRELIMINARY; PRT; 488 AA.
 AC Q96BB8
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein (similar to diacylglycerol O-acyltransferase
 DE homolog 1) (Mouse).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Skin;
 RA Strausberg R.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RA Strausberg R.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC015762; AAH15762.1; -;
 DR EMBL; BC023565; AAH23565.1; -;
 DR InterPro: IPR004299; MBOAT_fam.
 DR Pfam: PF03062; MBOAT; 1.
 KW Hypothetical protein; Acyltransferase; Transferase.
 SQ SEQUENCE 488 AA; 55278 MW; 6574D5DBF15D6171 CRC64;

Query Match 28.6%; Score 791.5; DB 4; Length 488;
 Best Local Similarity 38.0%; Pred. No. 3.8e-58;
 Matches 194; Conservative 77; Mismatches 174; Indels 65; Gaps 17;

RA Nylkiforuk C.L., Laroche A., Weselake R.J.;
 RT "Isolation and Sequence Analysis of a Novel cDNA Encoding a Putative
 RT Diacylglycerol Acyltransferase from a Microspora-derived Cell
 RT Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No.
 RT AF155224). (PGR99-123).";
 RL Plant Physiol. 120:1207-1207(1999).
 DR EMBL; AF155224; AAD40861.1; -;
 KW Acyltransferase; Transferase.
 SQ SEQUENCE 341 AA; 39532 MW; 1A46340C49F16332 CRC64;

Query Match 27.9%; Score 773.5; DB 11; Length 500;
 Best Local Similarity 37.5%; Pred. No. 1.3e-56;
 Matches 194; Conservative 86; Mismatches 167; Indels 71; Gaps 18;

Qy 14 ENGSGFEVLDLRLRRRSKRSDDSSNGL-LLSGSDNNPSDDVGAPADVRDRIDSVDNDQAQ 72
 Db 22 QGSGPMDVEEVN-DAAVGPDLAGGADAPAPA-PVPAPAHTRDK-----DRQ 67
 Qy 73 GTANLAGDNNGGDNNGGGGRGNADATFTYRPSVPAHRRARESPLSSDAIFKQSH 132

RESULT 9
 Q8BH15 PRELIMINARY; PRT; 500 AA.
 AC Q8BH15
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Diacylglycerol acyltransferase.
 GN DGAT.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Various strains;
 RA Harada Y., Watanabe T.K., Tanigami A.;
 RT "rat DGAT complete cds.";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB062759; BAC43739.1; -;
 DR EMBL; AB062760; BAC43740.1; -;
 DR EMBL; AB062761; BAC43741.1; -;
 DR EMBL; AB062762; BAC43742.1; -;
 DR EMBL; AB062763; BAC43743.1; -;
 KW Acyltransferase; Transferase.
 SQ SEQUENCE 500 AA; 57086 MW; 1A8592E7762AFFBE CRC64;

Query Match 27.9%; Score 773.5; DB 11; Length 500;
 Best Local Similarity 37.5%; Pred. No. 1.3e-56;
 Matches 194; Conservative 86; Mismatches 167; Indels 71; Gaps 18;

Qy 14 ENGSGFEVLDLRLRRRSKRSDDSSNGL-LLSGSDNNPSDDVGAPADVRDRIDSVDNDQAQ 72
 Db 22 QGSGPMDVEEVN-DAAVGPDLAGGADAPAPA-PVPAPAHTRDK-----DRQ 67
 Qy 73 GTANLAGDNNGGDNNGGGGRGNADATFTYRPSVPAHRRARESPLSSDAIFKQSH 132

```

OS Perilla frutescens.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiids; Lamiales; Lamiaceae; Nepetoideae; Elsholtziaceae;
OC Perilla.
OX NCBI_TaxID=48386;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Okdong;
RA Hwang S., Hwang Y.;
RT "Isolation of Perilla frutescens diacylglycerol acyltransferase
   RT cDNA";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF98815; AAG23696.1; -.
KW Acyltransferase; Transferase.
SQ SEQUENCE      534 AA;  61205 MW;  8FFC173E06E5BB70 CRC64;

Query Match          66.9%; Score 1854; DB 10; Length 534;
Best Local Similarity 65.3%; Pred. No. 1.9e-147;
Matches 359; Conservative 49; Mismatches 92; Indels 50; Gaps 9

Qy 1 MAILDSAGVTTVT----ENGGGFVFDLRLLRRKRSRSSNGLLLSGSDNNSPSDDVGAP 56
    ||||| : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 1 MAILDSEPIILDTTSSADNGAHHTTL--RRQSARSYPP---LLDSDSNSL----- 47

Qy 57 ADVEDRIDSVVND--DAQGTANLAGDNNG-----GDNNG-----GORGGE 96
    || : || : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 48 -----EAESAINDSENVRNDANLTENLRGAVESENEKEQSYGREGARVENGETSGN 102

Qy 97 GRGNADATFTVPSPVPARRARESPLESSDAIFPKOSHAGLENLCVVVLIAVNSRLIENLM 156
    || : || : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 103 GTDVMAVKFTFRPAAPARHNKNKEPLESSDAIFPKOSHAGLFNLICIVVLVAVNSRLIENLM 162

Qy 157 KYGWLIIRTDWFSSSRSLRDWPLMCCISLSIFPLAAFTVEKLVILOKYISEPVVFELHII 216
    || : || : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 163 KYGWLIKSGFWFSSTSLRDWPMLMCCLSLPVALASFLVEKLVLKLVIPPEWAVFLHVTI 222

Qy 217 TMTVELPYVYTILRCDSNFLSGVTIMLLCTCIWLKLVSYAHTSYDIRSIANAADK----- 271
    || : || : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 223 TTVELLPFYVVYLRCDSNAVLGSVGLTMFLFACVTKLVSYAHTNYDLRYLAKSLDWEAMS 282

Qy 272 --ANPEYSYYSLSKLAYFWAPTLCYCOPSPRASCKRGWWAROFAKLIVFTGPMGFIT 329
    || : || : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 283 RYNWLDYAYDVSFKSLAYFWAPTLCQPSYPTRACIKRGWWVRQIKLVIFTGLMGFIT 342

Qy 330 EQYNINPIVRNSKHPLKGDLLYAIERVLKSLPNLYVMCMFYCFPHLMNTLIAELLCRGD 389
    || : || : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 343 EQYNINPIVQNQHPLKGNLLYAIERVLKSLVPNLYVMCMFYCPFHLMNTLIAELLCRGD 402

Qy 390 REFYKDWNNAKSVGDYWRWMNMNVHKWNRHIYFPCLRKKTPKTALIIAFIVSAVFHEL 449
    ||||| : || : || : | : | : | : | : | : | : | : | : | : | : | : |
Db 403 REFYKDWNMNARTVEYWRWMNMNVHKWNRHIYPCQLONGIPKIYAVLIAFIIVSAIFAHEL 462

Qy 450 CIAYPCRLEFKLWAELIGTMFOVPLVFIETNYLQERF--GSTVGNNMFIFIQFGPCMVCVLLY 508
    || : || : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 463 CVAYPCQEIFRWAEFGSIMQLVPLIVTYNIQEKFNSMVGNMMFWFCFCIFGQPCMVCVLLY 522

Qy 509 YHDLMNKRGS 518
    ||||| || |
Db 523 YHDLMNKRAS 532

```

RESULT 6	
Q9SEG9	
ID	Q9SEG9
AC	PRELIMINARY; PRT; 532 AA.
DT	Q9SEG9;
DT	01-MAY-2000 (TREMBrel.. 13, Created)
DT	01-MAY-2000 (TREMBrel. 13, Last sequence update)
DT	01-MAR-2003 (TREMBrel. 23, Last annotation update)
DE	Diacylglycerol acylCOA acyltransferase.
GN	DGAT.
OS	Nicotiana tabacum (Common tobacco).
OC	Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC	Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
OX	NCBI_TaxID=4097;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=cv. xanthi SH6; TISSUE=Callus;
RX	MEDLINE=20069349; PubMed=10601854;
RA	Bouvier-Nave P., Benveniste P., Oelkers P., Sturley S.L., Schaller H.;
RT	"Expression in yeast and tobacco of plant cDNAs encoding acyl
RT	CoA:diacylglycerol acyltransferase.";
RL	Eur. J. Biochem. 267:85-96(2000).
RD	EMBL; AF129003; AAF19345.1; -
KW	- Acyltransferase; Transferase.
DW	
QY	SEQUENCE 532 AA; 60867 MW; C9D316E7A8799310 CRC64;

Query Match	66.5%;	Score	1842.5;	DB	10;	Length	532;
Best Local Similarity	67.0%;	Fred. No.	1.8e-104;				
Matches	358;	Conservative	49;	Mismatches	104;	Indels	23;
						Gaps	7;

QY	1	MAILDSAGVTVTWTEGGGEFVDLD---	RLRRKRSKSDSSGILLGSSDNNSSPDDVGAPA	57
Db	4	MELPESVEMTTTTTSGTENSLDNHVSRRR---	CGNGFEAASAINSS---DANMSE	56
QY	58	DVRDRIDS-----VYNDGDAOGTANLAGDNGGDNNGGGRGGEGEGRGNADAT----	FTYR	108
Db	57	DRRDVCGSGAGLETWNRKSKSGESSDVIKREDDRDNDNVANCEESKSTETTTTPEKFAYR	116	
QY	109	PSVPAHRRARSPSSDAIFQKSHAGLNLVWVLIIVNSRLIIENLMKYGHLITDFWF	168	
Db	117	ASAPAHRRIRKESPLSSDAIFQKSHAGLNLVWVLIIVNSRLIIENLMKYGILLIRAGFW	176	
QY	169	SSRSLRDWPMLCMCISLSIFPLAAFTVEKLVLOKVISEPVVIFLHIITMTVEVLPVYVT	228	
Db	177	SKSLRDWPLMLCCLSLQILPLAFLVEKLAQORHLTERAVVTLHITTTAAILFPVLVI	236	
QY	229	LRCDSAFSLGVTMLLLTIIWMLKLVSYAHTSYDIRSLANAADK---ANPEVSYYYSLKSL	285	
Db	237	LGCDSAFLFGVILMLVACIVWMLKLVSYAHTNHDMLQAKSTDQDTSDFSDVSKSL	296	
QY	286	AYFWVAPTLQVQSPYRSACIRKGVARQFAKLVITFTGFMGPIIIOYINPIVRNSKHPLK	345	
Db	297	AYFWVAPTLQVQSPYPTPCIRKGVARQFAKLVITFTGLMGFIIRQYINPIVQNSQHPLK	356	
QY	346	GDLVYATERVKLSPVNLVYVWLCMFYCFPHLWNLNLAEELCFGRFEFYKDMWNAKSGVDY	405	
Db	357	GNLLYAIERVKLSPVNLVYVWLCMFYCFPHLWNLNLAEELCFGRFEFYKDMWNAKTIDYEY	416	
QY	406	WRMNMMPVHKWVRHIYPPCLRSRKTPKPLAIITIAFLVSAPFHELCIAVPCRFLKWLAFUG	465	
Db	417	WRMNMMPVHKWVRHIYPPCLRNGLPKGVAIIIAFLVSAPFHELCIAVPCRFLKWWAFMG	476	
QY	466	IMFQVPLVIFINYLOERF-GSTVGNMIFWIFCFIGQPMCVLLYYHYDLNMRKGS	518	
Db	477	IMFQVPLVILTNFLQNKFOSSVMGNMFWCFECILGQPMCVLLYYHYDVMNRKSS	530	

RESULT	7
ID	Q9XGV4
AC	Q9XGV4;
DT	01-NOV-1999 (TREMBlrel. 12, Created)
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT	01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE	Putative diacylglycerol acyltransferase (EC 2.3.1.20).

RESULT 7
Q9XGV4
ID Q9X
AC Q9X
DT 01-
DT 01-
DT 01-
DE Put
GN DGA
OS Bra
OC Euk
OC Spe
OC eur
OX NCB
RN [1]
RP SEQ
RC STR

```

RESULT 3
Q9XGR5 PRELIMINARY; PRT; 503 AA.
AC Q9XGR5;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DE 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Putative diacylglycerol acyltransferase (EC 2.3.1.20).
GN DGAT1.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3708;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cultivar Jet Neuf;
RX MEDLINE=20027448; PubMed=10557255;
RA Nykiforuk C.L., Laroche A., Weselake R.J.;
RT "Isolation and Characterization of a cDNA Encoding a Second Putative
RT Diacylglycerol Acyltransferase from a Microspore-derived Cell
RT Suspension Culture of Brassica napus L. cv Jet Neuf (Accession No.
RT AF164434). (PGR99-158).";
RL Plant Physiol. 121:1053-1053(1999).
DR EMBL; AF164434; AAD45536.1; -.
KW Acyltransferase; Transferase.
SQ SEQUENCE 503 AA; 56931 MW; 2B578A16FE0AD758 CRC64;

Query Match 83.3%; Score 2309.5; DB 10; Length 503;
Best Local Similarity 84.8%; Pred. No. 9e186;
Matches 442; Conservative 17; Mismatches 43; Indels 19; Gaps 5;

QY 1 MAILDSAGVTT-VTENGGEEFVDLRLRRKRSRSDSSNGLLGGSDNNKSPDDVGPADV 59
DQ 1 MAILDSAGVAVPTENG---VADLRLRRKRSRSDSSNGLL---SDSPSDVGAARAE 53
60 RDRDSVNDAAQGTANLAGNNGGNGGGRGGEGGRGNADATFTYRPSVPAHRRARE 119
DQ 54 RDRVDSAAEEAQTANLA-----GGDAETRESAGG-----DVRFTYRPSVPAHRRTRE 102
120 SPLSSDAIFKQSHAGLNLVAVNSRLIIENLMKYGLWIRTDFFWESSRSLRDWPLF 179
DQ 103 SPLSSDAIFKQSHAGLNLVAVNSRLIIENLMKYGLWIRTDFFWESSRSLRDWPLF 162
180 MCCISLSIFPLAAFTVEKVLQKYESEPVVFLHIIITMTVEVLYPVVYTLRCDSAFLSGV 239
DQ 163 MCCISLSVFLAAFTVEKVLQKYESEPVVFLHIIITMTVEVLYPVVYTLRCDSAFLSGV 222
240 TMLLTCTVWLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPLICYOPS 299
DQ 223 TMLLTCTVWLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFMVAPLICYOPS 282
300 YPRSAIRKGVARQFAKLVIPTGFMGFIIBQYINPIVRNSKHPKLGDLVIAIERVLKLS 359
DQ 283 YPRSAIRKGVARQFAKLVIPTGFMGFIIBQYINPIVRNSKHPKLGDLVIAIERVLKLS 342
360 VPVLYVWLCMFYCFPHLWNLTAELLCFGDREFYKDMWNAKSVGDYWRMNNPVHKMVR 419
DQ 343 VPVLYVWLCMFYCFPHLWNLTAELLCFGDREFYKDMWNAKSVGDYWRMNNPVHKMVR 402
420 HIYFPCLSKIPKTLAIIIAFLVSAVPHLCAVPCRLFKLWAFGLMFPVPLFITNVL 479
DQ 403 HIYFPCLSKIPKTLAIIIAFLVSAVPHLCAVPCRLFKLWAFGLMFPVPLFITNVL 462
480 QERFGSTVGNMIFWIFCIGFQPCMCVLLYYHDLNMRKGSMS 520
DQ 463 QERFGSTVGNMIFWIFCIGFQPCMCVLLYYHDLNMRKGSMS 503

RESULT 4
Q8RX96 PRELIMINARY; PRT; 518 AA.
ID Q8RX96
AC Q8RX96;

```

```

DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Putative diacylglycerol acyltransferase.
DE Tropaeolum majus (Common nasturtium).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Tropaeolaceae; Tropaeolum.
OX NCBI_TaxID=4020;
RN [1]
RP SEQUENCE FROM N.A.
RA Mietkiewska E., Pedersen K., Katavic V., Taylor D.C.;
RT "Characterization of a putative diacylglycerol acyltransferase mRNA
RT from Tropaeolum majus embryo.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY084052; AAM03340.1; -.
DR InterPro; IPR004299; MBOAT.fam.
DR InterPro; IPR002040; Tachykinin.
DR Pfam; PF03062; MBOAT; 1.
DR PROSITE; PS00267; TACHYKININ; 1.
KW Transferase; Acyltransferase.
SQ SEQUENCE 518 AA; 58814 MW; 401D36427284CDA CRC64;

Query Match 67.8%; Score 1880; DB 10; Length 518;
Best Local Similarity 69.8%; Pred. No. 1.2e-149;
Matches 370; Conservative 44; Mismatches 88; Indels 28; Gaps 10;

QY 1 MAILDSAGVTTVTENGGEFVDLRLRRKRSRSD---SSNGLLGGSDNNKSPDDVGPADV 56
DQ 1 MAVAESQNTT-TWGGHGD-SDLNFRKPKSSSVIEPSSSGFT---STNGVPA---TGHV 53
57 ADVDRDRIDSVNDAAQGTANLAGNNGGNGGGRGGEGGR---GNADATFTYRPSVPAH 114
DQ 54 AENRQDQVRGAMENATGVSNLIG-----NGGGVVIGNEEKQVGETDIRTYRSPFAH 106
115 RRARESPLSSDAIFKQSHAGLNLVAVNSRLIIENLMKYGLWIRTDFFWESSRSLR 174
DQ 107 RRARESPLSSDAIFKQSHAGLNLVAVNSRLIIENLMKYGLWIRTDFFWESSRSLR 166
175 DWPLFMCCISLSIFPLAAFTVEKVLQKYESEPVVFLHIIITMTVEVLYPVVYTLRCDSA 234
DQ 167 DWPLFMCCISLSIFPLAAFTVEKVLQKYESEPVVFLHIIITMTVEVLYPVVYTLRCDSA 226
235 FLSGVTMLLTCVWLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAY 287
DQ 227 YMSGVWLMFCINWLVKLVSYAHTSSDIRTLAKSGYKGDHPNSIVSCSYDVSLSLAY 286
288 FMVAPLICYOPSYPASACIRKGVARQFAKLVIPTGFMGFIIEQYINPIVRNSKHPKLG 347
DQ 287 FMVAPLICYOPSYPASACIRKGVARQFAKLVIPTGFMGFIIEQYINPIVRNSKHPKLG 346
348 LLYAIERVLKLVSPNLYVWLCMFYCFPHLWNLTAELLCFGDREFYKDMWNAKSVGDYWR 407
DQ 347 LLYAIERVLKLVSPNLYVWLCMFYCFPHLWNLTAELLCFGDREFYKDMWNAKSVGDYWR 406
408 MNNMPVHKMVRHLYFPCLRSKIPKTLAIIIAFLVSAVPHLCAVPCRLFKLWAFGLM 467
DQ 407 MNNMPVHKMVRHLYFPCLRSKIPKTLAIIIAFLVSAVPHLCAVPCRLFKLWAFGLM 466
468 FOVPLVFTNTYLOERP-GSTVGNMIFWIFCIGFQPCMCVLLYYHDLNMRK 516
DQ 467 FOVPLVFTNTYLOERP-GSTVGNMIFWIFCIGFQPCMCVLLYYHDLNMRK 516

RESULT 5
Q9FUL6 PRELIMINARY; PRT; 534 AA.
ID Q9FUL6
AC Q9FUL6;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Diacylglycerol acyltransferase.
GN DGAT1.

```

RX MEDLINE-99313150; PubMed-10386579;
 RA Hills M.J., Lu C., Hobbs D.H.;
 RT "Cloning of a cDNA encoding diacylglycerol acyltransferase from
 RT Arabidopsis thaliana and its functional expression.";
 RL FEBS Lett. 452:145-149(1999).
 RN [5]
 RN SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
 RA Somerville C.R., Venter J.C.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RN SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Town C.D., Kaul S.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC005917; AD10144.2; -
 DR EMBL; AJ238008; CAB45373.1; -
 DR EMBL; AF051849; AAF19262.1; -
 DR EMBL; AJ131831; CAB44774.1; -
 DR EMBL; AY054480; AAK96671.1; -
 DR EMBL; AC003058; AAM14875.1; -
 KW Acyltransferase; Transferase.
 SQ SEQUENCE 520 AA; 58985 MW; 9CD0E3E8956CEFF4 CRC64;

Query Match 100.0%; Score 2771; DB 10; Length 520;
 Best Local Similarity 100.0%; Pred. No. 1.5e-224;
 Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAILDSAGVTVTENGSGGEFVDLRLRRKRSRSDSSNGLLSGDNNSSDDVGPADVR 60
 DB 1 MAILDSAGVTVTENGSGGEFVDLRLRRKRSRSDSSNGLLSGDNNSSDDVGPADVR 60
 QY 61 DRIDSVVNDDAOCTANLAGDNGGNGGGRGGEGGNADATFYRPSVPAHRRARES 120
 DB 61 DRIDSVVNDDAOCTANLAGDNGGNGGGRGGEGGNADATFYRPSVPAHRRARES 120
 QY 121 PLSSDAIFKQSHAGLENLCVVLIAVNSRLIENLMKYGLIRTFDFWSSRLRDWPLFM 180
 DB 121 PLSSDAIFKQSHAGLENLCVVLIAVNSRLIENLMKYGLIRTFDFWSSRLRDWPLFM 180
 QY 181 CCISLSIFPLAFTVEKLVQKVISEPVVIFLHIITMTVEVLPVVTLCDSAFLSGVT 240
 DB 181 CCISLSIFPLAFTVEKLVQKVISEPVVIFLHIITMTVEVLPVVTLCDSAFLSGVT 240
 QY 241 LMLLTCIVMLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFWVAPLTCYQPSY 300
 DB 241 LMLLTCIVMLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFWVAPLTCYQPSY 300
 QY 301 PRSACIRKGWARQFAKLVITFGFMGFIIEQYINPVRNSKHPKLGDLLEYAIEVRLKLSV 360
 DB 301 PRSACIRKGWARQFAKLVITFGFMGFIIEQYINPVRNSKHPKLGDLLEYAIEVRLKLSV 360
 QY 361 PNLYVWLCMFYCFHFLWLNLIAELLFCGDFREFYKDWNAKSVGDYWRMNMMPVHKWVRH 420
 DB 361 PNLYVWLCMFYCFHFLWLNLIAELLFCGDFREFYKDWNAKSVGDYWRMNMMPVHKWVRH 420
 QY 421 IYFPCILRSIKPTLAIITIAFLVSAVHELCIAVPCRLFKLWFLGIMFQVPLVFTINYLQ 480

DB 421 IYFPCILRSIKPTLAIITIAFLVSAVHELCIAVPCRLFKLWFLGIMFQVPLVFTINYLQ 480
 QY 481 ERFGSTVGNMIFWIFCIFGQPMCVLLYYHDLNMRKGSMS 520
 DB 481 ERFGSTVGNMIFWIFCIFGQPMCVLLYYHDLNMRKGSMS 520

RESULT 2
 Q9M4V2
 ID Q9M4V2 PRELIMINARY; PRT; 501 AA.
 AC Q9M4V2;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Putative diacylglycerol acyltransferase (EC 2.3.1.20).
 OS Brassica napus (Rape).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Brassica.
 OX NCBI_TaxID=3708;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Commanche;
 RA Brown A.P., Schierer T.P., Slabas A.R.;
 RT "Characterization of a putative diacylglycerol acyltransferase cDNA
 RT from Brassica napus embryo.";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF251794; AAF64065.1; -
 KW Acyltransferase; Transferase.
 SQ SEQUENCE 501 AA; 57538 MW; 29E022B278D60822 CRC64;

Query Match 84.9%; Score 2351.5; DB 10; Length 501;
 Best Local Similarity 86.0%; Pred. No. 2.6e-189;
 Matches 447; Conservative 17; Mismatches 37; Indels 19; Gaps 5;

QY 1 MAILDSAGVTVTENGSGGEFVDLRLRRKRSRSDSSNGLLSGDNNSSDDVGPADVR 60
 DB 1 MEILDGGVTMTPTENG---ADLTLRHRKRSRSDSSNGLL---PDSVTYSD-----ADVR 49
 QY 61 DRIDSVVNDDAOCTANLAGDNGGNGGGRGGEGGNADATFYRPSVPAHRRARES 120
 DB 50 DRVDSAV-EDTQCKANLAGENE-----IRESGAGGNGVDVRYRPSVPAHRRVRES 101
 QY 121 PLSSDAIFKQSHAGLENLCVVLIAVNSRLIENLMKYGLIRTFDFWSSRLRDWPLFM 180
 DB 102 PLSSDAIFKQSHAGLENLCVVLIAVNSRLIENLMKYGLIRTFDFWSSRLRDWPLFM 161
 QY 181 CCISLSIFPLAFTVEKLVQKVISEPVVIFLHIITMTVEVLPVVTLCDSAFLSGVT 240
 DB 162 CCISLSIFPLAFTVEKLVQKVISEPVVIFLHIITMTVEVLPVVTLCDSAFLSGVT 221
 QY 241 LMLLTCIVMLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFWVAPLTCYQPSY 300
 DB 222 LMLLTCIVMLKLVSYAHTSYDIRSLANAADKANPEVSYVSLKSLAYFWVAPLTCYQPSY 281
 QY 301 PRSACIRKGWARQFAKLVITFGFMGFIIEQYINPVRNSKHPKLGDLLEYAIEVRLKLSV 360
 DB 282 PRSPCIRKGWARQFAKLVITFGFMGFIIEQYINPVRNSKHPKLGDLLEYAIEVRLKLSV 341
 QY 361 PNLYVWLCMFYCFHFLWLNLIAELLFCGDFREFYKDWNAKSVGDYWRMNMMPVHKWVRH 420
 DB 342 PNLYVWLCMFYCFHFLWLNLIAELLFCGDFREFYKDWNAKSVGDYWRMNMMPVHKWVRH 401
 QY 421 IYFPCILRSIKPTLAIITIAFLVSAVHELCIAVPCRLFKLWFLGIMFQVPLVFTINYLQ 480
 DB 402 VYFPCILRRNIPKPAIILAFVSAVHELCIAVPCRLFKLWFLGIMFQVPLVFTINYLQ 461
 QY 481 ERFGSTVGNMIFWIFCIFGQPMCVLLYYHDLNMRKGSMS 520
 DB 462 ERFGSTVGNMIFWIFCIFGQPMCVLLYYHDLNMRKGSMS 501

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2003, 04:28:05 ; Search time 72 Seconds
(without alignments)
1863.713 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771

Sequence: 1 MAIDSAGVTVTENGGEF.....QPMCVLLYYHDLNMRKSGMS 520

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL_23:*
- 1: sp_archaea:*
 - 2: sp_bacteria:*
 - 3: sp_fungi:*
 - 4: sp_human:*
 - 5: sp_invertebrate:*
 - 6: sp_mammal:*
 - 7: sp_mmc:*
 - 8: sp_organelle:*
 - 9: sp_phase:*
 - 10: sp_plant:*
 - 11: sp_rodent:*
 - 12: sp_virus:*
 - 13: sp_vertebrate:*
 - 14: sp_unclassified:*
 - 15: sp_virus:*
 - 16: sp_bacteriap:*
 - 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	520	10 Q9SLD2	Q9sls2 arabidopsis
2	2351.5	84.9	501	10 Q9MAV2	Q9mav2 brassica na
3	2309.5	83.3	503	10 Q9XGR5	Q9xgr5 brassica na
4	1880	67.8	518	10 Q8RX96	Q8rx96 tropaeolum
5	1854	66.9	534	10 Q9FUL6	Q9ful6 perilla fru
6	1842.5	66.5	532	10 Q9SEG9	Q9seg9 nicotiana t
7	1717	62.0	341	10 Q9XGV4	Q9xgv4 brassica na
8	791.5	28.6	488	4 Q96BB8	Q96bb8 homo sapien
9	773.5	27.9	500	11 Q8BHI5	Q8bhi5 rattus norv
10	760.5	27.4	489	6 Q8SQB0	Q8sqb0 bos taurus
11	749.5	27.0	489	6 Q8MK44	Q8mk44 bos taurus
12	739.5	26.7	489	6 Q8WHZ1	Q8whz1 sus scrofa
13	735.5	26.5	403	4 Q9BRH5	Q9brh5 homo sapien
14	725.5	26.2	565	5 Q96OU8	Q96ou8 drosophila
15	724.5	26.1	496	5 Q45245	Q45245 caenorhabdi
16	724.5	26.1	498	5 Q9NCE1	Q9nce1 caenorhabdi

17	720.5	26.0	565	5 Q8ST50	Q8st50 drosophila
18	557	20.1	653	5 Q97295	Q97295 plasmodium
19	539	19.5	242	11 Q91YB5	Q91yb5 rattus norv
20	478	17.3	401	5 Q8SR71	Q8sr71 encephalito
21	419.5	15.1	462	11 Q8C795	Q8c795 mus musculu
22	397	14.3	508	6 Q8WNN5	Q8wnn5 gorilla gor
23	397	14.3	550	4 Q8NIE4	Q8nie4 homo sapien
24	390.5	14.1	508	6 Q8WNN4	Q8wnn4 pongo pygma
25	378.5	13.7	525	11 Q8R0Y9	Q8r0y9 mus musculu
26	374	13.5	559	5 Q8MYW9	Q8myw9 drosophila
27	357.5	12.9	522	4 Q96TD4	Q96td4 homo sapien
28	350.5	12.6	524	11 Q8K1W9	Q8klm9 rattus norv
29	340.5	12.3	292	5 Q9VHN2	Q9vhn2 drosophila
30	339	12.2	580	3 Q9C2E9	Q9c2e9 neurospora
31	329	11.9	472	3 Q9U0U8	Q9uu82 schizosacch
32	279	10.1	467	5 Q17498	Q17498 caenorhabdi
33	165.5	6.0	305	6 Q95214	Q95214 oryctolagus
34	162	5.8	609	3 Q08929	Q08929 saccharomyc
35	149.5	5.4	455	16 Q8XK89	Q8xk89 clostridium
36	147.5	5.3	621	10 Q9LNR9	Q9lnk9 arabidopsis
37	144.5	5.2	404	16 Q8CPW1	Q8cpw1 staphylococ
38	142	5.1	458	16 Q9PH05	Q9phg5 campylobact
39	139.5	5.0	404	16 Q53662	Q53662 staphylococ
40	134.5	4.9	865	16 Q8Y619	Q8y619 listeria mo
41	127	4.6	510	16 Q8F6E3	Q8f6e3 leptospira
42	126	4.5	495	16 Q83577	Q83577 treponema p
43	122.5	4.4	473	16 Q8F4B7	Q8f4b7 leptospira
44	121.5	4.4	450	16 Q926V2	Q926v2 chlamydia p
45	121.5	4.4	1170	15 Q9PNZ5	Q9pnz5 campylobact

ALIGNMENTS

RESULT 1

Q9SLD2 ID Q9SLD2 PRELIMINARY; PRT: 520 AA.

AC Q9SLD2; Q9S7F2;
 DT 01-MAY-2000 (Tremblrel. 13, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Diacylglycerol O-acyltransferase (Diacylglycerol
 DE acyltransferase).
 GN AT2G19450 OR DGAT OR DGAT OR F3P11.5.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,
 RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,
 RA Carrera A.J., Creasy T.H., Buell C.R., Town C.D., Nierman W.C.,
 RA Fraser C.M., Venter J.C.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Zou J., Wei Y., Jiao C., Selvaraj G., Taylor D.C.;
 RT "The Arabidopsis thaliana TAG1 gene encodes for a diacylglycerol
 RT acyltransferase.";
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE-Hypocotyl;
 RX MEDLINE=20069349; PubMed=10601854;
 RA Bouvier-Nave P., Benveniste P., Oelkers P., Sturley S.L., Schaller H.;
 RT "Expression in yeast and tobacco of plant cDNAs encoding acyl
 RT COA:diacylglycerol acyltransferase.";
 RL Eur. J. Biochem. 267:85-96(2000).
 RN [4]
 RP SEQUENCE FROM N.A.

Cercopithecinae; Cercopithecus.
 NCBI_TaxID=9534;
 [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=98434590; PubMed=9756918;
 RA Anderson R.A., Joyce C., Davis M., Reagan J.W., Clark M.,
 RA Shelness G.S., Rudel L.L.;
 RT "Identification of a form of acyl-CoA:cholesterol acyltransferase
 RL J. Biol. Chem. 273:26747-26754(1998).
 CC !- FUNCTION: PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY
 CC CHOLESTEROL ABSORPTION. IN ADDITION TO ITS ACYLTRANSFERASE
 CC ACTIVITY, IT MAY ACT AS A LIGASE. MAY PROVIDE CHOLESTERYL ESTERS
 CC FOR LIPOPROTEIN SECRETION FROM HEPATOCYTES AND INTESTINAL MUCOSA.
 CC !- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol
 CC ester.
 CC !- SUBUNIT: MAY FORM HOMO- OR HETERODIMERS (BY SIMILARITY).
 CC !- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum.
 CC !- TISSUE SPECIFICITY: LIVER AND INTESTINE. EXPRESSION IS SIX-FOLD
 CC GREATER IN HEPATOCYTES THAN IN KUPFFER CELLS.
 CC !- DISEASE: ACCUMULATION OF INSOLUBLE CHOLESTEROL ESTERS IN
 CC MACROPHAGES AND SMOOTH MUSCLE IS A CHARACTERISTIC FEATURE OF EARLY
 CC LESIONS OF ATHEROSCLEROTIC PLAQUE.
 CC !- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF053234; AAC62929.1; --
 CC Pfam; PF03062; MBOAT; 1.
 CC Transferrase; Acyltransferase; Transmembrane; Endoplasmic reticulum;
 CC Cholesterol metabolism.
 CC FT TRANSMEM 124 144 POTENTIAL.
 CC FT TRANSMEM 155 175 POTENTIAL.
 CC FT TRANSMEM 200 220 POTENTIAL.
 CC FT TRANSMEM 308 328 POTENTIAL.
 CC FT TRANSMEM 340 360 POTENTIAL.
 CC FT TRANSMEM 416 436 POTENTIAL.
 CC FT TRANSMEM 441 461 POTENTIAL.
 CC FT TRANSMEM 476 496 POTENTIAL.
 CC SEQUENCE 526 AA; 60411 MW; F45E29F48386DB43 CRC64;
 Query Match 12.6%; Score 350.5; DB 1; Length 526;
 Best Local Similarity 26.8%; Pred. No. 5.1e-19;
 Matches 114; Conservative 69; Mismatches 156; Indels 87; Gaps 17;
 QY 132 HAGLENCVVVLIIVNSRLIENLMKYGLWLRITDFWSSRSRLRDPWPLFMCISISIPPLA 191
 DB 125 HMFAGLCVFI--STLAI-----DFIDGRLLLEFDLLI--FSFGQLPLA 166
 QY 192 AFTVEKLVQYISEPVVIFLHIITTEVL-----YPVVTLRCDSAFLSGVT 240
 DB 167 LVTV-----VPMFLSTLLAPYQALRLNARPCARTWILGAGLGC--ALLAAHA 212
 QY 241 LMLIT-----CIVMLKLVSAHTSYD--IRSLANAADKAN-----PEV 276
 DB 213 LVLCALPVAHVAEHQLPASRCVLVEQVRLMKSYSLREAVPGTLKARRGEGIQAPSF 272
 QY 277 SYVYSLKSLAYEMVAPILCQPSYPSRNCIRKGVARQFAKLVITFGFMGFIHQYINPI 336
 DB 273 SSYL-----YFLFCPTLIYRETPRTPIYRWNVAKNFAQALGCVLYACFILGLCVPV 326
 QY 337 VRN-SKHPKGLDLYATERVKLSPNLYVWLMFYCFEHLWLNILAEILCFGRFVKD 395
 DB 327 FANMSREPFSTRAL--VLSILHLATPGIFMILLIFLFAFLHCLWLNFAEMLRFGDMFYRD 384

396 WNAKSVGDYWRMNMMPVHKWVRHIYFP--CLRSKIPKTLAIIAFLVSAVHELCIAV 453
 DB 385 WNSTSFNSYRTNWNVVDWLYSYVQDGLWLLGAARGVAMLGVLVSAVAHEYIPC- 443
 QY 454 PCRLFKLWAFGLGIMFOYPLVF--ITNY-LQRFSGSTVGNMIFWIFCFIGQPMCVLLYYH 510
 DB 444 ----FVLGFFVPVILFLVIGGMLNFMHMDQHTGPANWLVMTML-FLGOGIOVLYCQ 498
 QY 511 DLMNRK 516
 DB 499 EWIYAR 504

RESULT 15
 AREL_YEAST
 ID AREL_YEAST STANDARD; PRT; 610 AA.
 AC P25628;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Sterol-ester synthase 1).
 GN AREL OR SAT2 OR YCR048W OR YCR48W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 CC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Grivell L.A., de Haan M., Maat M.J.;
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE OF 1-328 FROM N.A.
 RA Bolotin-Fukuhara M., Buhler J.-M., Daignan-Fornier B., Doira C.,
 RA Francques-Gaillard M.-C.;
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
 [3]
 RP CHARACTERIZATION.
 RX MEDLINE=96243137; PubMed=8650549;
 RA Yang H., Bard M., Bruner D.A., Gleeson A., Deckelbaum R.J.,
 RA Aljinovic G., Pohl T.M., Rothstein R., Sturley S.L.;
 RL "Sterol esterification in yeast: a two-gene process";
 RL Science 272:1353-1356(1996).
 CC !- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol
 CC ester.
 CC !- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum.
 CC !- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X59720; CAA42296.1; --
 CC PIR; S19461; S19461.
 CC SGD; S0000644; AREL.
 CC GO; GO:0005783; C:endoplasmic reticulum; IDA.
 CC GO; GO:0004772; F:sterol O-acyltransferase activity; IDA.
 CC GO; GO:0016125; P:sterol metabolism; IMP.
 CC Pfam; PF03062; MBOAT; 1.
 CC Transferrase; Transmembrane; Endoplasmic reticulum; Acyltransferase.
 CC FT TRANSMEM 182 202 POTENTIAL.
 CC FT TRANSMEM 229 249 POTENTIAL.
 CC FT TRANSMEM 264 284 POTENTIAL.
 CC FT TRANSMEM 371 391 POTENTIAL.
 CC FT TRANSMEM 409 429 POTENTIAL.
 CC FT TRANSMEM 535 555 POTENTIAL.
 CC FT TRANSMEM 590 610 POTENTIAL.
 CC SEQUENCE 610 AA; 71613 MW; FF72EFB9238B2205 CRC64;

Db 417 LGRRARGVAMLGVELVSAVVEHYIFC-----FVLGFFVPMVLMFLVFGGLNFTMDNRH 471

QY 484 GSTVGNMIFWFCIFGQPMCVLLYYHDLNMRK 516

Db 472 TGPANWILMW-TFLPMGOGIOVSLYCOEWYARR 503

RESULT 13

SOA2_HUMAN

ID SOA2_HUMAN STANDARD; PRT; 522 AA.

AC 075908; Q9UNR2;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Sterol O-acyltransferase 2 (EC 2.3.1.26) (Cholesterol acyltransferase 2) (Acyl coenzyme A:cholesterol acyltransferase 2) (ACAT-2).

DE 2) (Acyl coenzyme A:cholesterol acyltransferase 2) (ACAT-2).

GN SOAT2 OR ACAT2 OR ACAT2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RX SEQUENCE FROM N.A.

RX MEDLINE=98434592; PubMed=9756920;

RA Oelkers P., Behari A., Cromley D., Billheimer J.T., Sturley S.L.;

RT "Characterization of two human genes encoding acyl coenzyme

RT A:cholesterol acyltransferase-related enzymes.";

RL J. Biol. Chem. 273:26765-26771(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Intestine;

RX MEDLINE=20428724; PubMed=10846185;

RA Chang C.C.Y., Sakashita N., Ornvold K., Lee O., Chang E.T., Dong R.,

RA Lin S., Lee C.Y.G., Strom S.C., Kashyap R., Fung J.J.,

RA Farese R.V. Jr., Patoisseau J.-F., Delhon A., Chang T.-Y.;

RT "Immunological quantitation and localization of ACAT-1 and ACAT-2 in

RT human liver and small intestine.";

RL J. Biol. Chem. 275:28083-28092(2000).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=21225021; PubMed=11325614;

RA Katsuren K., Tamura T., Arashiro R., Takata K., Matsuura T.,

RA Nikiawa N., Ohta T.;

RT "Structure of the human acyl-CoA:cholesterol acyltransferase-2

RT (ACAT-2) gene and its relation to dyslipidemia.";

RL Biochim. Biophys. Acta 1531:230-240(2001).

CC -!- FUNCTION: PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY

CC CHOLESTEROL ABSORPTION. IN ADDITION TO ITS ACYLTRANSFERASE

CC ACTIVITY, IT MAY ACT AS A LIGASE. MAY PROVIDE CHOLESTERYL ESTERS

CC FOR LIPOPROTEIN SECRETION FROM HEPATOCYTES AND INTESTINAL MUCOSA.

CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol - CoA + cholesterol

CC ester.

CC -!- SUBUNIT: MAY FORM HOMO- OR HETERODIMERS (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic

CC reticulum.

CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

DR EMBL; AF059203; AAC63998.1; -

DR EMBL; AF099031; AAC78335.2; -

DR EMBL; AF331516; AAK18275.1; -

DR EMBL; AF331502; AAK18275.1; JOINED.

DR EMBL; AF331503; AAK18275.1; JOINED.

DR EMBL; AF331504; AAK18275.1; JOINED.

DR EMBL; AF331505; AAK18275.1; JOINED.

DR EMBL; AF331506; AAK18275.1; JOINED.

DR EMBL; AF331507; AAK18275.1; JOINED.

DR EMBL; AF331508; AAK18275.1; JOINED.

DR EMBL; AF331509; AAK18275.1; JOINED.

DR EMBL; AF331510; AAK18275.1; JOINED.

DR EMBL; AF331511; AAK18275.1; JOINED.

DR EMBL; AF331512; AAK18275.1; JOINED.

DR EMBL; AF331513; AAK18275.1; JOINED.

DR EMBL; AF331514; AAK18275.1; JOINED.

DR EMBL; AF331515; AAK18275.1; JOINED.

DR Genew; HGNC:11178; SOAT2.

DR MIM; 601311; -

DR GO; GO:0008415; F:acyltransferase activity; TAS.

DR GO; GO:0008203; P:cholesterol metabolism; TAS.

DR Pfam; PF03062; MBOAT; 1.

KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;

KW Cholesterol metabolism.

FT TRANSMEM 124 144 POTENTIAL.

FT TRANSMEM 155 175 POTENTIAL.

FT TRANSMEM 200 220 POTENTIAL.

FT TRANSMEM 262 282 POTENTIAL.

FT TRANSMEM 304 324 POTENTIAL.

FT TRANSMEM 344 366 POTENTIAL.

FT TRANSMEM 437 457 POTENTIAL.

FT TRANSMEM 472 492 POTENTIAL.

FT CONFLICT 254 254 T -> I (IN REF. 2).

SQ SEQUENCE 522 AA; 59896 MW; EEAC2DB569FF729 CRC64;

Query Match 12.9%; Score 357.5; DB 1; Length 522;

Best Local Similarity 27.8%; Pred. No. 1.5e-19;

Matches 115; Conservative 66; Mismatches 166; Indels 67; Gaps 16;

QY 132 HAGLENLCVVLIAVNSRLIETNLMKYGLWLTDFWFSRSRLDNPWFMCCLISLIFPLA 191

Db 135 HMFIAGLCVFII-----STLAI-----DFIDGRLLLEFDLLI--FSGQLPLA 166

QY 192 AFTVEKLVQKYISEPVVIFLIIITMTVEVLYPVVTVTLRCDLSAFSLGVTMLL----- 244

Db 167 LVTWVPFSLTLLAPYQALRLWARGTQA-----TGLGALLAAHAAVLCALPVHVAVE 221

QY 245 -----TCIVWLKLVSYAHTSYD--IRSLANAADKAN-----PEVSYSYSLKSLAYFM 289

Db 222 HOLPPASRCVLVFEQVRFMLKSYSPLEAVPGTLRARRGEGIQAPSFSSYL-----VFL 275

QY 290 VAPTLCYQPSYPSRACIRKGGWVARQFALVIFTGFMGTIIQYINPIYRN--SKHPLKGD 348

Db 276 FCPTLIYRETPRTPTPVVRNMYAKNFAQALGCVLYACFILTCLQVFPVFNAMSREPFSRA 335

QY 349 LYAIEERVLKLSVPLNYVLMCFYCFHILNLNLIAELLCGDREFYKDWNAKSVGDYWRM 408

Db 336 L--VLSILHATLPGIFMLLLIFFAFLHCLWLNFAEMLRFGDMFYRDWNNSTFSNYYRT 393

QY 409 WNPVHKWVRHIYPPCLR--SKIPKTLAIIAIFLVSVAHFELCTAVPCRUFKLMWAFGI 466

Db 394 WNVVVDWLYSVYQDGLRLLGARARGVAMLGVLVLSVAHVEYIFC-----FVLGFFFPV 448

QY 467 MEQVPLVF--ITNVL--QERGSTVGNMIFWFIIFGFCGPMCVLLYYHDLNMRK 516

Db 449 MLIDFVGGLNFMHMDORTG--PAWNLMWMTML--FLGGGIQVSLYCOEWYARR 500

RESULT 14

SOA2_CERAE

ID SOA2_CERAE STANDARD; PRT; 526 AA.

AC 077759;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Sterol O-acyltransferase 2 (EC 2.3.1.26) (Cholesterol acyltransferase 2) (Acyl coenzyme A:cholesterol acyltransferase 2) (ACAT-2).

GN SOAT2 OR ACAT2.

OS Cercopithecus aethiops (Green monkey) (Grivet).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;

DR GO: GO:0004772; F:sterol O-acyltransferase activity; IDA.
 DR GO:0016125; P:sterol metabolism; IMP.
 KW Pfam; PF03062; MBOAT; 1.
 FT TRANSFERASE: Transmembrane; Endoplasmic reticulum; Acyltransferase.
 FT TRANSMEM 215 235
 FT TRANSMEM 292 312
 FT TRANSMEM 404 424
 FT TRANSMEM 442 462
 FT TRANSMEM 485 505
 FT TRANSMEM 567 587
 FT TRANSMEM 622 642
 FT CONFLICT 80 80
 FT CONFLICT 184 184
 FT CONFLICT 211 211
 FT CONFLICT 612 612
 FT CONFLICT 642 642
 FT SEQUENCE 642 AA; 74022 MW; 035FC4ED9C7CD830 CRC64;
 Query Match 13.78; Score 378.5; DB 1; Length 642;
 Best Local Similarity 25.28; Pred. No. 5.1e-21;
 Matches 133; Conservative 83; Mismatches 176; Indels 135; Gaps 19;
 QY 102 DAYFTYRPSV-----PAHRRARE-----SPLSSDAIF 128
 DB 125 DMSFEHRPSILDSVNEPFTKVGPLEKEIRREKELAMKRLNHRKSSPDVDSVG 184
 QY 129 KSHAGLFNLCV-----VLAVNSRLIEN---LMKYGWL-----IRTFWFS-SRS 172
 DB 185 KNDGAAPTTPVTAATSETVTVETIISNFSGLYVAFVMAIAFGAVKALIDYYQHNGS 244
 QY 173 LRQWPLMCCISLIFPLAFTVEKLVQKYSIEPVVIFLH-----IIIT 217
 DB 245 FKDEI-----LKFWTNLTVASVDLLMYLTFYVVGIOYLCKGVLKWTGTGWTFS 298
 QY 218 MTEVLYPVVYTLRDSA-----FLSGVTMLLTCTIVWLKLYSA-----HTSYDI 262
 DB 299 IYEFELFVIFMYLTENILKHLNLSKIFLHLSLVLLMKHSFAYNGYLWGIRKELOFSK 358
 QY 263 RSLANAADKAN-PEV-----SYVSLKSLAYPMVAPT 294
 DB 359 SALAKYKDSINDPKVIGALEKSCFCSFELSSQSLSQTKQFPNNISAKSFFWTMPPTL 418
 QY 295 CYQPSYPRSAICIRKGWARQFAKIVITGPMGFIIQYI-NPI-----VRNSKHLKGD 347
 DB 419 IYQLEYPRTEIRRSYVLEKIC-AIFGTIFLMDIAQILMPVAMRALAVRNS-----E 471
 QY 348 LLXAIERVLKL-----SVNPLY-VLCMFYCFHMLNITLAEILLCFGDREFFYKDMWNAK 400
 DB 472 WTGILDELLKFWGLLDIVGFIWYILDYFLWDAILNCVAELTRGDRYFFYGDWNCV 531
 QY 401 SVGDYWRMNNPVHKWVRHYPFCLRS-KIPKTLAIITIAFLVSAVPELHICIAVPCRLFK 459
 DB 532 SWADFSRIWNPVHKFLRHVHSSMSFKNLSQATLMTFFLSSVVELAMVYIFKKLR 591
 QY 460 LWAFGLTMFQVPLVFTNYLQERPGSVGNMIFWFCIFGQPCNVL 506
 DB 592 FYLFFQMLQPLVALNTKFMNRRTIIGNVIFWLGICMGPSVNCNL 638

RESULT 12

SOA2_MOUSE
 ID SOA2_MOUSE STANDARD; PRT; 525 AA.
 AC O88908;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sterol O-acyltransferase 2 (EC 2.3.1.26) (cholesterol acyltransferase
 2) (Acyl coenzyme A:cholesterol acyltransferase 2) (ACAT-2).
 GN SOAT2 OR ACAT2
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RX MEDLINE=98434591; PubMed=9756919;
 RA Cases S., Novak S., Zheng Y.-W., Myers H.M., Lear S.R., Sande E.,
 RA Welch C.B., Lusis A.J., Spencer T.A., Krause B.R., Erickson S.K.,
 RA Farese R.V. Jr.;
 RT "ACAT-2, a second mammalian acyl-CoA:cholesterol acyltransferase. Its
 RT cloning, expression, and characterization.";
 RL J. Biol. Chem. 273:26755-26764(1998).
 CC -1- FUNCTION: PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY
 CC CHOLESTEROL ABSORPTION. IN ADDITION TO ITS ACYLTRANSFERASE
 CC ACTIVITY, IT MAY ACT AS A LIGASE. MAY PROVIDE CHOLESTERYL ESTERS
 CC FOR LIPOPROTEIN SECRETION FROM HEPATOCYTES AND INTESTINAL MUCOSA.
 CC -1- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol -> CoA + cholesterol
 CC ester.
 CC -1- SUBUNIT: MAY FORM HOMO- OR HETERODIMERS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum.
 CC -1- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AF078751; AAC64057.1;
 DR MGD; MGI:1332226; Soat2.
 DR Pfam; PF03062; MBOAT; 1.
 KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;
 KW Cholesterol metabolism.
 FT TRANSMEM 123 143 POTENTIAL.
 FT TRANSMEM 154 174 POTENTIAL.
 FT TRANSMEM 197 217 POTENTIAL.
 FT TRANSMEM 265 285 POTENTIAL.
 FT TRANSMEM 307 327 POTENTIAL.
 FT TRANSMEM 350 370 POTENTIAL.
 FT TRANSMEM 440 460 POTENTIAL.
 FT TRANSMEM 475 495 POTENTIAL.
 FT SEQUENCE 525 AA; 60823 MW; 01CDSZE206B5C397 CRC64;
 Query Match 13.5%; Score 375; DB 1; Length 525;
 Best Local Similarity 27.4%; Pred. No. 7.4e-21;
 Matches 124; Conservative 73; Mismatches 162; Indels 94; Gaps 19;
 QY 112 PAHRR---ARESPSSDAIFKQSHAGLFNLCVAVNSRLIENLMKYGLIRTFDWF 168
 DB 97 PGRKKVFVARKSLT--DELMEVQH---FRTIYHMFIAGLWFLIISTL-----AIDFID 144
 QY 169 SSRSLRDWPLMCCISLSIFPLAFTVEKLVQKYSIEPVVIFLHIITTEVLYPVVVT 228
 DB 145 EGRMLLEFDLL--FSFGQLPALMTW-----VPMFLYTLVVPYQTLW-LMAR 189
 QY 229 LRCDSAFLSGVTL--MLLT-----CIVWLKLVSYAHTSYD- 261
 DB 190 PRAGGAMWLGASLGCVLAAHVVLCVLPVHVSVRHELPPASRCVLYFEQVRLMKSYSF 249
 QY 262 -----IRSLANAADRANPEVSYVSLKSLAYPMVAPTLCYQPSYPRSAICIRKGWARQ 314
 DB 250 LRETVPGIFCVRRKGISPPSFSSYL-----YFLCPTLIYREYRTPRSIRWNYVAKN 303
 QY 315 FAKIVITGPMGFIIQYIINPIVEN-SKHPKLGDLLEYAIEKLVLSV-----PNLYWLC 368
 DB 304 FAQVGLCLLYACFTILGRLCVFVFANMSREPPS-----TRALLSLILHATGPGIFMLL 356
 QY 369 MFYCFHMLNITLAEILLCFGDREFFYKDMWNAKSYGDMWNNMNVHKKWVRHYPF--CL 426
 DB 357 IFFAFLHCLWLNAPFAELRFGDRMFPYRDWNSTSFSTNYRTNWNVVDHLSYVYQDGLWL 416
 QY 427 RSKIPKTLAIITIAFLVSAVPELHICIAVPCRLFKLWAFGLTMFQVPLVFTNYLQERPGSVGNMIFWFCIFGQPCNVL 483

Db 468 LSEFFYVLFVLEFFGMAFN-----FIVNDSRKK---PIWVLMWTSIFLNGVLLCFYSQ 520

RESULT 10

SOAL_RAT

ID SOAL_RAT STANDARD; PRT; 545 AA.

AC 070536;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase 1) (ACAT-1).

DE 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).

GN SOAT1 OR ACAT OR ACAT.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.

OX NCBI_TaxID=10116;

[1]

RP SEQUENCE FROM N.A.

RC STRAIN-Wistar; TISSUE-Adrenal gland;

RX MEDLINE=98223432; PubMed=955010;

RA Matsuda H., Hakamata H., Kawasaki T., Sakashita N., Miyazaki A., Takahashi K., Shichiri M., Horiuchi S.;

RT "Molecular cloning, functional expression and tissue distribution of rat acyl-coenzyme A:cholesterol acyltransferase.";

RL Blochim. Biophys. Acta 1391:193-203(1998).

CC -|- FUNCTION: CATALYZES THE FORMATION OF FATTY ACID-CHOLESTEROL ESTERS. PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY CHOLESTEROL ABSORPTION.

CC -|- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol ester.

CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum.

CC -|- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC

DR EMBL; D86373; BAA25372.1; -

DR Pfam; PF03062; MB0AT; 1.

KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;

KW Cholesterol metabolism.

FT TRANSMEM 136 156 POTENTIAL.

FT TRANSMEM 316 336 POTENTIAL.

FT TRANSMEM 357 377 POTENTIAL.

FT TRANSMEM 465 485 POTENTIAL.

FT TRANSMEM 493 513 POTENTIAL.

SQ SEQUENCE 545 AA; 64146 MW; 40129EF21257BBF CRC64;

Query Match 14.1%; Score 391.5; DB 1; Length 545;

Best Local Similarity 28.0%; Pred. No. 4.4e-22;

Matches 132; Conservative 78; Mismatches 151; Indels 111; Gaps 24;

QY 99 GNADATFTYRPSVPAHRRARE--SPLSS-----DAIFKQSH-AGLFNLCVWVLI 144

Db 86 GCALTFTSILEEMKNHRAKDLRAPPEQKIFISRLSDELFVDHRTIYMFALLI 145

QY 145 AV-----NSRLITE-NLMKYGV-----LIRDFWSSRSRLDPLMCCISLSI 187

Db 146 IFILSTLVVDYIDEGRLVLEFSLAYAFGFPPIVWT--W-----WAMFLSTLAIPY 195

QY 188 FPLAATFVKLVQKI-----SEPVVIFL-----HIITMTEVL--YPPVYTLRCDRAF 235

Db 196 F-----LQFRAHYSKSSHPLIYSLINGAFPLVQLGILGFIPTTYVYVYATLTPP 245

QY 236 LSGVTLMLTCTIVLWKLVSAYHTSYDIRSLANAADKAN-----PEVSYVYSLKSLAYFWA 291

Db 246 ASRFILILEQIRLVMAHSYVRENVP-RVLSAKESSTVPTVQVYL-----YFLFA 298

QY 292 PILCYOPSPRSACIRKKGWVARQFAKLVIITFGMGFIIEQYINPIVRNSK-HPLKGDLLY 350

Db 299 PTLIYRDSYPRPTPTRWGVAMQFLQVFGCLFVYVIFERLCAPLFRNFKQBPESA--- 354

QY 351 ATERVLKLSV-----PNLYVWLCMFYCFHNLWNLIAELLFCGDRFYEKDMNAKSVGDY 405

Db 355 ---RVVLVCVFNILPGVLMFLSFFAFHLCWNLNFAEMLRFGDMFYKDMWNSTSYNS 411

QY 406 WRWNNPVPKHWVRHIYFCL--RSKIPKTLAIIAFLVSASFHELCIAVPCR-----L 457

Db 412 YRTWNVVHDWLYVYVYKDLLNFFSRFRPAAMLAVFALSVAVVEYALAV-CLSYFPVVL 470

QY 458 FKLMAELGIMFQVPLVFIITNYLQERGSTVGNNI-----FWFTFCIFGQ 501

Db 471 FVLEMFPGMAFN-----FIVNDSRKR---PVWIMVRASLFLGHGVILCFYSQ 515

RESULT 11

ARE2_YEAST

ID ARE2_YEAST STANDARD; PRT; 642 AA.

AC P53629; Q12673;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Sterol O-acyltransferase 2 (EC 2.3.1.26) (Sterol-ester synthase 2).

GN ARE2 OR SAT1 OR YNR019W OR N3206.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RA Pohl T.M.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A., AND CHARACTERIZATION.

RX MEDLINE=96243137; PubMed=8650549;

RA Yang H., Bard M., Bruner D.A., Gleeson A., Deckelbaum R.J., Aljinovic G., Pohl T.M., Rothstein R., Sturley S.L.;

RT "Sterol esterification in yeast: a two-gene process.";

RL Science 272:1353-1356(1996).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=SNY243;

RX MEDLINE=96394550; PubMed=8798656;

RA Yu C., Kennedy N.J., Chang C.C.Y., Rothblatt J.A.;

RT "Molecular cloning and characterization of two isoforms of Saccharomyces cerevisiae acyl-CoA:sterol acyltransferase.";

RL J. Biol. Chem. 271:24157-24163(1996).

CC -|- FUNCTION: ENSURES PROBABLY MOST OF THE ACYLTRANSFERASE ACTIVITY. SUPPRESSION OF ARE2 REDUCES STEROL ESTER LEVELS TO 25% OF THE NORMAL VALUE.

CC -|- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol ester.

CC -|- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum.

CC -|- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

CC

DR EMBL; 271634; CAA96298.1; -

DR EMBL; U51790; AAB02203.1; -

DR EMBL; U55383; AAC49441.1; -

DR PIR; S63350; S63350.

DR SGD; S0005302; ARE2.

DR GO; GO:0005783; C:endoplasmic reticulum; IDA.

QY 71 AGGTANLAGDNNGGNGGGRGGGGRGNADATYTPSPVAHRAESPSSDAIFKQ 130
 DB 118 -QGKIFI-----ARRSLDELLEVDHIRT 141
 QY 131 SHAGFNLCVVVLIIV-----NSRLIIE-NLMKYG-LIRDFWFSRSLRDWPLFMC 182
 DB 142 YHMFALLILFILSTLVVDYIDEGRLVLEFSLLSYAFGKFPVW-----TWIMF 192
 QY 183 ISLSIFPLAAFTVEKLVQKYI-----SEPVV--IFLHIITMTEV-----LYPVVTLR 230
 DB 193 -----LSTFSPYFLQRMATGYSKSHPLNSLFGFLFVQIGILGPGYVVL- 244
 QY 231 CDSAPLGSVTLMLLTCIVWLKLVSAHT---SYDIRSLANAADKAN---PEVSYVSLK 283
 DB 245 ---AYTLPPASRFIIIFEQIRFVMAKHSFVRENPRVLSNAKESKSTVPITPVNYL- 298
 QY 284 SLAYFWAPTLCYQPSRACIRKGWVARQAKLVIETGFGFIIEQYINPIVNSK-H 342
 DB 299 ---YFLFAPTLYRDSYPRNPTVRMGYVAMQAFVQFCFFVYVYIFERLCAPLFRNIKE 355
 QY 343 PLKGDLLYAIEIRVLKLSV-----PNLYVWLCMFYCFHLLNLIAELLCFGDREYKDW 397
 DB 356 PFSA-----RVLVLCVFNILPGVILILFTFFAFHLCHWLNFAEMLRFGDRMEYKDW 408
 QY 398 NAKSGDYWRMNMMPVHKWVRHIY--FPCLRSKIPKTLAIIAFVLSAVFHELCAIYPC 455
 DB 409 NSTSYSNYRTYRNVVVDWLYYAYKDFLWLFPSKRSAAMLAVPAVSAVVEYALAV-C 467
 QY 456 R-----LFKLWAFIGIMFQVPLVITNYLQERFGSTVGNMIFW-----FIFCIGQ 501
 DB 468 LSFFYPVLVLFVFMFFGMAFN-----FIVNDSRKK-----PIWNVNMWTSLSFLGNGVLICFY 520

RESULT 9

SOAL_HUMAN
 ID SOAL_HUMAN STANDARD; PRT: 550 AA.
 AC P35610;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase
 DE 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).
 GN SOAT1 OR SOAT OR STAT OR ACAT1 OR ACAT2 OR ACAT1 OR ACAT.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Macrophage;
 RX MEDLINE=94012607; PubMed=8407899;
 RA Chang C.-Y., Huh H.-Y., Cadigan K.M., Chang T.-Y.;
 RT "Molecular cloning and functional expression of human acyl-coenzyme
 RT A:cholesterol acyltransferase cDNA in mutant Chinese hamster ovary
 RT cells.";
 RL J. Biol. Chem. 268:20747-20755(1993).
 RN [2]
 RP REVISION TO 207.
 RA Chang C.-Y., Chang T.-Y.;
 RL Submitted (MAY-1999) to the SWISS-PROT data bank.
 RN [3]
 RP TOPOLOGY.
 RX MEDLINE=99367457; PubMed=10438503;
 RA Lin S., Cheng D., Liu M.S., Chen J., Chang T.-Y.;
 RT "Human acyl-CoA:cholesterol acyltransferase-1 in the endoplasmic
 RT reticulum contains seven transmembrane domains.";
 RL J. Biol. Chem. 274:23276-23285(1999).
 CC -!- FUNCTION: CATALYZES THE FORMATION OF FATTY ACID-CHOLESTEROL
 CC ESTERS. PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY
 CC CHOLESTEROL ABSORPTION. IN ADDITION TO ITS ACYLTRANSFERASE
 CC ACTIVITY, IT MAY ACT AS A LIGASE.
 CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol -> CoA + cholesterol

ester.
 -!- SUBUNIT: MAY FORM HOMO- OR HETERODIMERS.
 -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum.
 CC -!- INDUCTION: HIGHLY ACTIVATED BY THE PRESENCE OF CHOLESTEROL.
 CC -!- DISEASE: ACCUMULATION OF INSOLUBLE CHOLESTEROL ESTERS IN
 CC MACROPHAGES AND SMOOTH MUSCLE IS A CHARACTERISTIC FEATURE OF
 CC EARLY LESIONS OF ATHEROSCLEROTIC PLAQUE.
 CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L21934; AAC37532.2; -
 DR PIR; A59038; A48026.
 DR Genew; HGNC:11177; SOAT1.
 DR MIM; 102642; -
 DR GO; GO:0005783; C:endoplasmic reticulum; TAS.
 DR GO; GO:0008415; F:acyltransferase activity; TAS.
 DR GO; GO:0008203; P:cholesterol metabolism; TAS.
 DR Pfam; PF03062; MBOAT; 1.
 KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;
 KW Cholesterol metabolism.
 FT TRANSMEM 141 159 POTENTIAL.
 FT TRANSMEM 320 341 POTENTIAL.
 FT TRANSMEM 361 382 POTENTIAL.
 FT TRANSMEM 470 490 POTENTIAL.
 FT TRANSMEM 498 518 POTENTIAL.
 SQ SEQUENCE 550 AA; 64762 MW; 5C6AFE525D541DEE CRC64;

Query Match 14.3%; Score 397; DB 1; Length 550;
 Best Local Similarity 26.7%; Pred. No. 1.7e-22;
 Matches 144; Conservative 75; Mismatches 177; Indels 144; Gaps 24;
 QY 17 GGEFVDLRLRRKSRSSNGL-----LLSGSDNNSPSDVGPADVRDRIDSVND 70
 DB 70 GSHFDFTVNLIERASLDNGCALTFVSLEGEKNHRKDLRAPPE----- 117
 QY 71 AQGTANLAGDNNGGNGGGRGGGGRGNADATYTPSPVAHRAESPSSDAIFKQ 130
 DB 118 -QGKIFI-----ARRSLDELLEVDHIRT 141
 QY 131 SHAGFNLCVVVLIIV-----NSRLIIE-NLMKYG-LIRDFWFSRSLRDWPLFMC 182
 DB 142 YHMFALLILFILSTLVVDYIDEGRLVLEFSLLSYAFGKFPVW-----TWIMF 192
 QY 183 ISLSIFPLAAFTVEKLVQKYI-----SEPVV--IFLHIITMTEVLY-PYVVTIR 230
 DB 193 -----LSTFSPYFLQRMATGYSKSHPLNSLFGFLFVQIGILGPGYVVL- 244
 QY 231 CDSAPLGSVTLMLLTCIVWLKLVSAHT---SYDIRSLANAADKAN---PEVSYVSLK 283
 DB 245 ---AYTLPPASRFIIIFEQIRFVMAKHSFVRENPRVLSNAKESKSTVPITPVNYL- 298
 QY 284 SLAYFWAPTLCYQPSRACIRKGWVARQAKLVIETGFGFIIEQYINPIVNSK-H 342
 DB 299 ---YFLFAPTLYRDSYPRNPTVRMGYVAMQAFVQFCFFVYVYIFERLCAPLFRNIKE 355
 QY 343 PLKGDLLYAIEIRVLKLSV-----PNLYVWLCMFYCFHLLNLIAELLCFGDREYKDW 397
 DB 356 PFSA-----RVLVLCVFNILPGVILILFTFFAFHLCHWLNFAEMLRFGDRMEYKDW 408
 QY 398 NAKSGDYWRMNMMPVHKWVRHIY--FPCLRSKIPKTLAIIAFVLSAVFHELCAIYPC 455
 DB 409 NSTSYSNYRTYRNVVVDWLYYAYKDFLWLFPSKRSAAMLAVPAVSAVVEYALAV-C 467
 QY 456 R-----LFKLWAFIGIMFQVPLVITNYLQERFGSTVGNMIFW-----FIFCIGQ 501

OC Cercopithecinae; Macaca.
 RN NCBI_Taxid=9541;
 RP SEQUENCE FROM N.A.
 RC TISSUE-Adrenal gland;
 RX MEDLINE=98434590; PubMed=9756918;
 RA Anderson R.A., Joyce C., Davis M., Reagan J.W., Clark M.,
 RA Shelness G.S., Rudel L.L.;
 RT "Identification of a form of acyl-CoA:cholesterol acyltransferase
 specific to liver and intestine in nonhuman primates.";
 RL J. Biol. Chem. 273:26747-26754(1998).
 CC -!- FUNCTION: PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY
 CC CHOLESTEROL ABSORPTION. IN ADDITION TO ITS ACYLTRANSFERASE
 CC ACTIVITY, IT MAY ACT AS A LIGASE (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol
 CC ester.
 CC -!- SUBUNIT: MAY FORM HOMO- OR HETERODIMERS (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, BUT MOST STRONGLY
 CC IN THE ADRENAL GLAND. EXPRESSED MORE STRONGLY IN LIVER KUPFFER
 CC CELLS THAN IN HEPATOCYTES.
 CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF053337; AAC62931.1;
 DR Pfam; PF03062; MBOAT; 1.
 KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;
 KW Cholesterol metabolism.
 FT TRANSMEM 142 162 POTENTIAL.
 FT TRANSMEM 321 341 POTENTIAL.
 FT TRANSMEM 362 382 POTENTIAL.
 FT TRANSMEM 470 490 POTENTIAL.
 FT TRANSMEM 498 518 POTENTIAL.
 FT TRANSMEM 550 AA; 64699 MW; AA655206D58C291D CRC64;
 SQ SEQUENCE 550 AA; 64699 MW; AA655206D58C291D CRC64;
 Query Match 14.7%; Score 406; DB 1; Length 550;
 Best Local Similarity 26.7%; Pred. No. 3.7e-23;
 Matches 144; Conservative 76; Mismatches 176; Indels 144; Gaps 24;

QY 17 GGEFVDLRLRRKSRDSSNGL-----LLSGSDNNSPSDVGPADYRDRIDSVVND 70
 DB 70 GSHFDDFTVNLIKESASLDNGGCGALTTFSLGKNNHRAKDLRAPPE----- 117
 QY 71 AQGTANLAGDNNGGDNGGGRGGEGGCGNADATFTYRPSVPAHRRARESPSSDAIFRQ 130
 DB 118 -QKIFI-----ARRSLDELLEVDHIRT 141
 QY 131 SHAGLENLCVVLIIV-----NSRLIIE-NLMKYGW-LIRTDWFSSRLDNDWLFMC 182
 DB 142 YHMFIALLLFTLSTLVVDYIDEGRLVLEFSLSYAFGRFTVW-----TWIMF--- 192
 QY 183 ISLSIFPLAAFTVKEVLQKVI-----SEPVV--IFLHIITMTBV-----LVPVVTLR 230
 DB 193 -----LSTFSPVFLFORWATGYSKSHPLNSLFGHFLPMVFOIGILGPGTYVWL- 244
 QY 231 CDSAFUSGVTLLMTCIVWLKLVSYAHT---SYDIRSLANAADKAN-----PEVSYVSLK 283
 DB 245 ---AVTLPPASRFIIIFEQIREFVMAKHSFVRENVPRVLSAKESKSTVPIPTVQYL--- 298
 QY 284 SLAYFWVATLCYQSPRSACIRKGVARQAKLVIFGPMGFIEQVNPVIRNSK-H 342
 DB 299 ---YFLFAPTLIYRSDPRNPVWGVYAMQAFQVGEFCFFYYIIPERLCAPLFRNIQOE 355
 QY 343 PLKGDLLYAIVRLKLSV-----PNLYVWLCHFCFFHLMLNLIELLCGDFREFYKDW 397
 DB 70 GSHFDDFTVNLIKESASLDNGGCGALTTFSLGKNNHRAKDLRAPPE----- 117

DB 356 PFSA-----RVLVLCVENSILPGVLILFTFFAFLHCLWNAFAEMLRFGDRMFYKDW 408
 QY 398 NAKSVGDYWRMNMPPVHKWVRHIY--FPLRSKIPKTLAIITIAFVSAVFEHLCIAPVC 455
 DB 409 NSTSYSTYRTWNVVVDWLYYAYKDFLWFSKRFKSAMLAAFAVSVAWEYALAV-C 467
 QY 456 R-----LFKLWAFGLGIMFQVPLVFTNYLQBRFGSTVGNMIFW-----FIFCIFO 501
 DB 468 LSFFPVLVLFVLFMFGMAFN---FTVNDSRKK---PIWNVMWMTSLFLGNGVLLLCFYSQ 520
 RESULT 8
 SOAL_CERAE STANDARD; PRT; 550 AA.
 AC 077760;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase
 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).
 OS SOAT1 OR ACAT1.
 GN Cercopithecus aethiops (Green monkey) (Grivet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
 OC Cercopithecinae; Cercopithecus.
 OC NCBI_Taxid=9534;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC TISSUE-Adrenal gland;
 RX MEDLINE=98434590; PubMed=9756918;
 RA Anderson R.A., Joyce C., Davis M., Reagan J.W., Clark M.,
 RA Shelness G.S., Rudel L.L.;
 RT "Identification of a form of acyl-CoA:cholesterol acyltransferase
 specific to liver and intestine in nonhuman primates.";
 RL J. Biol. Chem. 273:26747-26754(1998).
 CC -!- FUNCTION: CATALYZES THE FORMATION OF FATTY ACID-CHOLESTEROL
 CC ESTERS. PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY
 CC CHOLESTEROL ABSORPTION.
 CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol
 CC ester.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum.
 CC -!- TISSUE SPECIFICITY: EXPRESSED IN MOST TISSUES, BUT MOST STRONGLY
 CC IN THE ADRENAL GLAND. EXPRESSED MORE STRONGLY IN LIVER KUPFFER
 CC CELLS THAN IN HEPATOCYTES.
 CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF053336; AAC62930.1;
 DR Pfam; PF03062; MBOAT; 1.
 KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;
 KW Cholesterol metabolism.
 FT TRANSMEM 142 162 POTENTIAL.
 FT TRANSMEM 321 341 POTENTIAL.
 FT TRANSMEM 362 382 POTENTIAL.
 FT TRANSMEM 470 490 POTENTIAL.
 FT TRANSMEM 498 518 POTENTIAL.
 FT TRANSMEM 550 AA; 64727 MW; 3B5E4CF8DB6CC713 CRC64;
 SQ SEQUENCE 550 AA; 64727 MW; 3B5E4CF8DB6CC713 CRC64;
 Query Match 14.5%; Score 402; DB 1; Length 550;
 Best Local Similarity 26.5%; Pred. No. 7.3e-23;
 Matches 143; Conservative 76; Mismatches 177; Indels 144; Gaps 24;

QY 17 GGEFVDLRLRRKSRDSSNGL-----LLSGSDNNSPSDVGPADYRDRIDSVVND 70
 DB 70 GSHFDDFTVNLIKESASLDNGGCGALTTFSLGKNNHRAKDLRAPPE----- 117

FT TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 311 331 POTENTIAL.
 FT TRANSMEM 352 372 POTENTIAL.
 FT TRANSMEM 460 480 POTENTIAL.
 FT TRANSMEM 488 508 POTENTIAL.
 FT CONFLICT 195 195 P -> R (IN REF. 2).
 SQ SEQUENCE 540 AA; 63739 MW; 8F900C8BCDF73C0 CRC64;
 Query Match 15.0%; Score 415.5; DB 1; Length 540;
 Best Local Similarity 28.5%; Pred. NO. 6.9e-24;
 Matches 135; Conservative 76; Mismatches 149; Indels 113; Gaps 24;

QY 99 GNADATFYRPSVAHRA-----RESPLSDAIFKQSHAG----- 134
 DB 81 GCALTTFISLEEMKNHRAKOLRAPPEQKGFISRSQSL--DELFEVDHIRTTHYHMFAL 138
 QY 135 --LNLVLLVLIAN--SRLIIE-NLMKYGW-LIRTFWFSSRLSDWPLFMCCTSLSTFP 189
 DB 139 LILFVLSIVVDYIDEGRLVLEFNLLAYAFKGFPTVW-----TWAMFLSTUSIPYF- 191
 QY 190 LAFTVEKLVLOKY-----SEPVVIFL-HIITMTEVL-----YPIVVTLCDSAPLS 237
 DB 192 -----LFOPMHAGYSKSSHPLIYSLVHGLLFLVFGVIGFVPTVVL-----AYTL 238
 QY 238 GVTMLLTCIYWLKLVSAHT--SYDIRSLANAADKAN-----PEVSYVYSLKSLAYPMV 290
 DB 239 PPASRFILILEQILIMKAHSFVRENIPRVLNAAKESKSDPLPTVNOYL-----YPLF 292
 QY 291 APTLCYOPSPRSACIRKGWARQFAKLVITFTGFMGFIIEQYINPIVNSK-HPLKGDLL 349
 DB 293 APTLIYRNYPTPTVRWGYYVAMQLQVFGCLFYVYIFERLCAPLFRNIKQEPESA--- 349
 QY 350 YAIERVLKLSV-----PNLYVWLCMYCFHFLMLNLAELLCFGDREFYKDMWNAKSYGD 404
 DB 350 ----RVLVLCVFNLSILPGVILILSLFFAFLHCLWLNAPAEMLRFGDRMFYKDMWNSTSYN 405
 QY 405 YWRMNPVHKWVRHIYFPCL--RSKIPKTLAIIIAFLVSAVHELCIAVPCR----- 456
 DB 406 YIRTNVNVHDWLYYYAYKDLLWFSSRFRKSAAMLAVALFSAVHEYALAI-CLSYFYPV 464
 QY 457 LFLKWAFLGIMFQVPLVFTNYLQERFGSTVGNMIFW-----FIFCIFGQ 501
 DB 465 LFLVLEFGMAFN---FIVNDSRKR---PIWIMVWASLFLGLILCLFYSQ 510

RESULT 6

SOAL_CRIGR
 ID SOAL_CRIGR STANDARD; PRT; 546 AA.
 AC Q60457;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).
 DE 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).
 GN SOAT1 OR ACAT1 OR ACAT.
 OS Cricetus griseus (Chinese hamster).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 CC Cricetulus.
 OX NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96278939; PubMed=8662991;
 RA Cao G., Goldstein J.L., Brown M.S.;
 RT "Complementation of mutation in acyl-CoA:cholesterol acyltransferase (ACAT) fails to restore sterol regulation in ACAT-defective sterol-resistant hamster cells."
 RL J. Biol. Chem. 271:14642-14648(1996).
 CC -!- FUNCTION: CATALYZES THE FORMATION OF FATTY ACID-CHOLESTEROL ESTERS. PLAYS A ROLE IN LIPOPROTEIN ASSEMBLY AND DIETARY CHOLESTEROL ABSORPTION.
 CC -!- CATALYTIC ACTIVITY: Acyl-CoA + cholesterol = CoA + cholesterol ester.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum.
 CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U47320; AAC52670.1; -;
 DR Pfam; PF03062; MBOAT; 1.
 KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum;
 KW Cholesterol metabolism.
 FT TRANSMEM 138 158 POTENTIAL.
 FT TRANSMEM 317 337 POTENTIAL.
 FT TRANSMEM 358 378 POTENTIAL.
 FT TRANSMEM 466 486 POTENTIAL.
 FT TRANSMEM 494 514 POTENTIAL.
 SQ SEQUENCE 546 AA; 64110 MW; 4ED6C403AAC7E65B CRC64;
 Query Match 14.9%; Score 412.5; DB 1; Length 546;
 Best Local Similarity 28.2%; Pred. NO. 1.2e-23;
 Matches 133; Conservative 78; Mismatches 151; Indels 109; Gaps 24;

QY 99 GNADATFYRPSVAHRAARE--SPLSS-----DAIFKQSH-----AGLFN 137
 DB 87 GCALTTFISLEEMKNHRAKOLRAPPEKGFISRRSLDELFEVDHIRTTHYHMFILILI 146
 QY 138 LCVVLIAN-----SRLIIE-NLMKYGW-LIRTFWFSSRLSDWPLFMCCTSLSTFP 191
 DB 147 LFILSTLVVDYIDEGRLVLEFNLLAYAFKGLPTVW-----TWAMFLSTUSIPYF--- 197
 QY 192 AFTVEKLVLOXY-----ISEPVVIFL-----HIITMTEVL--YPIVVTLCDSAPLSGV 239
 DB 198 -----LFOPMHAGYSKTSHPLIYSLSHGFLLVFGVIGFVPTVVL-----AYTLPP 246
 QY 240 TMLLTCIYWLKLVSAHT--SYDIRSLANAADKAN-----PEVSYVYSLKSLAYFMAP 292
 DB 247 ASRFIVILEQIRMYHKAHSFVRENIPRVLNAAKESSTVPTVNOYL-----YFLFAP 300
 QY 293 TLCYOPSPRSACIRKGWARQFAKLVITFTGFMGFIIEQYINPIVNSK-HPLKGDLLYA 351
 DB 301 TLIYRDSYPTPTVRWGYYVAMQLQVFGCLFYVYIFERLCAPLFRNIKQEPESA----- 355
 QY 352 IERVLKLSV-----PNLYVWLCMYCFHFLMLNLAELLCFGDREFYKDMWNAKSYGDY 406
 DB 356 --RVLVLCVFNLSILPGVILMLFLTFFAFLHCLWLNAPAEMLRFGDRMFYKDMWNSTSYN 413
 QY 407 RWMNPVHKWVRHIYFPCL--RSKIPKTLAIIIAFLVSAVHELCIAVPCR-----LF 458
 DB 414 RTNVNVHDWLYYYAYKDLLWFSSRFRKSAAMLAVALFSAVHEYALAV-CLSYFYPVLV 472
 QY 459 KLWAFGLGIMFQVPLVFTNYLQERFGSTVGNMIFW-----FIFCIFGQ 501
 DB 473 VLEFFFGMAFN---FIVNDSRKR---PIWIMVWASLFLGLHGLVILCLFYSQ 516

RESULT 7
 SOAL_MACFA
 ID SOAL_MACFA STANDARD; PRT; 550 AA.
 AC Q77761;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Sterol O-acyltransferase 1 (EC 2.3.1.26) (Cholesterol acyltransferase 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).
 DE 1) (Acyl coenzyme A:cholesterol acyltransferase 1) (ACAT-1).
 GN SOAT1 OR ACAT1 OR ACAT.
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;

DE acyltransferase).

GN DGAT1 OR DGAT.

OS Mus musculus (Mouse).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J;

RX MEDLINE=99007259; PubMed=9789033;

RA Cases S., Smith S.J., Zheng Y.-W., Myers H.M., Lear S.R., Sande E.,

RA Novak S., Collins C., Welch C.B., Lusis A.J., Erickson S.K.,

RA Farese R.V. Jr.;

RT Identification of a gene encoding an acyl CoA:diacylglycerol

RT acyltransferase, a key enzyme in triacylglycerol synthesis.*;

RL Proc. Natl. Acad. Sci. U.S.A. 95:13018-13023(1998).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=Tongue;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,

RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Gustincich S., Hill D., Hofmann C., Hume D.A., Kamiya M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuiki S.,

RA Hayashizaki Y.;

RT *Functional annotation of a full-length mouse cdna collection.*;

RL Nature 409:685-690(2001).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=22388257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A., Rubin G.M., Hong L.,

RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Udell T.B., Toshyuki S., Carninci P., Prange C.,

RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT *Generation and initial analysis of more than 15,000 full-length

RT human and mouse cdna sequences.*;

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

CC -1- FUNCTION: Catalyzes the terminal and only committed step in

CC triacylglycerol synthesis by using diacylglycerol and fatty acyl

CC CoA as substrates

CC -1- CATALYTIC ACTIVITY: Acyl-CoA + 1,2-diacylglycerol -> CoA +

CC triacylglycerol.

CC -1- PATHWAY: Central role in the metabolism of cellular diacylglycerol

CC lipids.

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic

CC reticulum.

CC -1- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF078752; AAC72917.1;

DR EMBL; AK008995; -; NOT_ANNOTATED_CDS.

DR EMBL; BC003717; AA03717.1; -;

DR MGD; MGI:1333825; Dgat1.

DR Pfam; PF03062; MBOAT; 1.

KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum.

FT TRANSMEM 113 133 POTENTIAL.

FT TRANSMEM 139 159 POTENTIAL.

FT TRANSMEM 175 195 POTENTIAL.

FT TRANSMEM 198 218 POTENTIAL.

FT TRANSMEM 293 313 POTENTIAL.

FT TRANSMEM 343 363 POTENTIAL.

FT TRANSMEM 417 437 POTENTIAL.

FT TRANSMEM 439 459 POTENTIAL.

FT TRANSMEM 464 484 POTENTIAL.

SO SEQUENCE 498 AA; 56790 MW; E7B0DD6DDCFIEC2B CRC64;

Query Match 27.8%; Score 771.5; DB 1; Length 498;

Best Local Similarity 36.3%; Pred. No. 1.2e-50;

Matches 190; Conservative 84; Mismatches 178; Indels 71; Gaps 15;

QY 8 GVTVTENGGEFVDLRLRRKRSRSDSSNGLLSGDNNSPSDVGVAPADVRDRIDSV 67

DB 16 GSRVSVGGSGPKVEDEVDAVSPD-----LCAGGDAPAPAP---APAHRDK----- 62

QY 68 NDDAQGTANLAGDNNGGGCGGEGRGGNADATFTTPSPVAHRARRRESPLSDAI 127

DB 63 -----DGRTSVGDGWDLR-----HRLQDSLFSSDSG 90

QY 128 FKOSHAGLNLVVVLIIVANSRLIENLMKVGILRTDFWFSRSLRD---WPLFMCCIS 184

DB 91 F-SNYRGILNWCVVVLLISNARLENLKYGILV-DPIQVVSFLKDPYSPAPCVIIA 148

QY 185 LSIFPLAAFTVEKLVQKYSIEPVVIFLHIITMEVLYPVVYVTLRCDSAFLSGVTMLL 244

DB 149 SNIFWAAFOIEKRLAVGALTEQMGLLHVVNLATIIICFPAAVALLVESITPVGSVFALA 208

QY 245 T-CIVWLKLYSYAHTSY-----DIRSLANAADK-----ANPEYSY--YVSLKSLAYPMV 290

DB 209 SYSIMFLKLYSYRDVNLWCRRVRKAKAVSTGKRVSGAAQAQVSPDNLTIRDLTYFIF 268

QY 291 APTLCYQSPSRACIRKGVVARQFAKLVITFTGFMFTIEQYINPIVRNSKHPKLGULLY 350

DB 269 APTLCYELNPSRPIRKRLRLRVLEMLFTQLQVGLIQWVPTTONSKMKPKF-DMDY 327

QY 351 A--IERVKLKSVNLYVWLCMFCYFFHLMNLIAELLCFGDREYFKDWNNAKSYGDWYM 408

DB 328 SRIIERLLKLVNPHLIWLIFFYWFHSCNLNAVAELQFGDREYFDWNNAESVYTFWQ 387

QY 409 WNPVHKWVRHYFPCLRSKIPKTLAIITAFLYSAVFHELCAVPCRLKFLWFLGIMP 468

DB 388 WNPVHKWVRHYFPCLRSKIPKTLAIITAFLYSAVFHELCAVPCRLKFLWFLGIMP 468

QY 469 QVPLVFTNTVQLQERFGSTVGNMIFWIFCFIQGPMCVLLVYVD 511

DB 448 QVPLAWIVGRF---FQGNYNAAVW-VTLIGQPVAVLMVYVD 486

RESULT 4

DGTL RAT ID DGTL RAT

AC O9PRM3; STANDARD; PRT; 498 AA.

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

FT TRANSMEM 453 473 POTENTIAL.
SQ SEQUENCE 488 AA; 55252 MW; 12E34BA7478ABAF CRC64;

Query Match 28.6%; Score 791.5; DB 1; Length 488;
Best Local Similarity 38.0%; Pred. No. 3.7e-52;
Matches 194; Conservative 77; Mismatches 174; Indels 65; Gaps 17;

QY 25 RLRRKRSRSDSSNGLLSGSDNNPSDDVAGADVPDRIDSVVNDDAQG-TANLAGDNG 83
  - - - - -
Db 8 RRRRTGSRPSSHG - - - - -GGPAAAE - - - - -VRDAAAGPDVGAAGDAFA 48

QY 84 GDNNGGGGGGGRGNADATFTYRSPVPAHRRARESPSSDAIFKQSHAGFLNLCVVVL 143
  - - - - -
Db 49 PAPKDDAGVSGHWELRC - - - - -HRLQSLFSSDSGF-SNTRGILNWCVVML 96

QY 144 IAVNSRLIENLMKYGLIRTDWFSSRLD - - - - -WPLFMCCISLFIPLAFTVEKLV 200
  - - - - -
Db 97 ILSNARLELENLIKYGILV-DPIQVVSFLKOPHSPAPCLVIAANVFAVAFQVEKRLA 155

QY 201 QKYSIEPVVIFLHIITMTVEVLPVVTLCDSAFSLGVTMLLT-CIWLKLVSAHTS 259
  - - - - -
Db 156 VQALTEQAGLLHVNATILCFPAAVLLVESITPVGSLALMAHTILFLKLSYR - - - 212

QY 260 YDTRS - - - - -LANADKANP - - - - -EVS - - - - -YVLSKSLAYFMVAPTLQPSYPSRA 304
  - - - - -
Db 213 -DVNSWCRARAKAASAGKASSAAAPHTVSPDNLTLYRDLVYFLFAPILCYELNPPSP 271

QY 305 CIRKGWVAFQAKLVITFGFMGFIIEQYINPIVRNSKHLKGLDLYA - - - - -IERVILKSVPN 362
  - - - - -
Db 272 RIRKRELLRLEMLFTQVLQGLIQOMVPTIOMSKPKF-DMDYSRIERLLKLVN 330

QY 363 LYVWLCMFCFFHLNLIALLCFDREYKDWNAKSVGDYWRWNNPVHKWVRHIY 422
  - - - - -
Db 331 HLWIFFYFLHSCILNAELMQLFGDREYKDWNSSEVTYFWQNNIPVHKWCIHFY 390

QY 423 FPLCRSKIPKTLAIITAFVSAVFHELCTAVPCRLKFWAFLGIMFQVPLV-FITVLOE 481
  - - - - -
Db 391 KPLMRGSKWMAKTGVFLASAFFHELYVSPRLMFLWAFMGMAQIPLANEVGRFFQ 450

QY 482 RFGSTVGNMIFWIFCIGQPCMCVLLYYHD 511
  - - - - -
Db 451 NY - - - - -GNAAVW-LSLIIGQPIAVLMYVHD 475

RESULT 2
DGT1_CERAE
ID DGT1_CERAE STANDARD; PRT; 491 AA.
AC Q9GME1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Diacylglycerol O-acyltransferase 1 (EC 2.3.1.20) (Diglyceride acyltransferase).
DE DGAT1 OR DGAT.
OS Cercopithecus aethiops (Green monkey) (Grivet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Joyce C.W., Davis M.A., Anderson R.A., Rudel L.L.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Catalyzes the terminal and only committed step in triacylglycerol synthesis by using diacylglycerol and fatty acyl CoA as substrates.
CC -!- CATALYTIC ACTIVITY: Acyl-CoA + 1,2-diacylglycerol -> CoA + triacylglycerol.
CC -!- PATHWAY: Central role in the metabolism of cellular diacylglycerol lipids.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum (By similarity).

```

```

CC -!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; AF236018; MBOAT5.1;
DR Pfam; PF03062; MBOAT5.1;
KW Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum.
FT TRANSMEM 107 127
FT TRANSMEM 131 151 POTENTIAL.
FT TRANSMEM 169 189 POTENTIAL.
FT TRANSMEM 192 212 POTENTIAL.
FT TRANSMEM 285 305 POTENTIAL.
FT TRANSMEM 335 355 POTENTIAL.
FT TRANSMEM 404 424 POTENTIAL.
FT TRANSMEM 431 451 POTENTIAL.
FT TRANSMEM 456 476 POTENTIAL.
SQ SEQUENCE 491 AA; 55643 MW; BFD3683453D588DB CRC64;

Query Match 27.9%; Score 772.5; DB 1; Length 491;
Best Local Similarity 37.8%; Pred. No. 9.9e-51;
Matches 191; Conservative 80; Mismatches 179; Indels 55; Gaps 17;

QY 25 RLRRKRSRSDSSNGLLSGSDNNPSDDVAGADVPDRIDSVVNDDAQG-TANLAGDNGS 84
  - - - - -
Db 11 RRRRTGSRPSSHG - - - - -GGPAAAE - - - - -VRDAAAG - - - - -DNGAA 46

QY 85 GDNNGGGGGGGRGNADATFTYRSPVPAHRRARESPSSDAIFKQSHAGFLNLCVVVL 144
  - - - - -
Db 47 GDAPAPAPSKDADGVDVAGHWELR - - - - -CH-RLQDSLFSSDSGF-NNYRGILNWCVVML 100

QY 145 AVNSRLIENLMKYGLIRTDWFSSRLD - - - - -WPLFMCCISLFIPLAFTVEKLV 201
  - - - - -
Db 101 LSNARLELENLIKYGILV-DPIQVVSFLKOPHSPAPCLVIAANVFAVAFQVEKRLAV 159

QY 202 KYISIEPVVIFLHIITMTVEVLPVVTLCDSAFSLGVTMLLT-CIWLKLVSAHTS 260
  - - - - -
Db 160 GALTEQAGLLHVNATILCFPAAVLLVESITPVGSLALMAHTILFLKLSYRDNVL 219

QY 261 DIR - - - - -SIANAADKANP - - - - -EVS - - - - -YVLSKSLAYFMVAPTLQPSYPSRACIRKG 309
  - - - - -
Db 220 WCRARAKAASAGKASSAAAPHTVSPDNLTLYRDLVYFLFAPILCYELNPPSPRIRKR 279

QY 310 WVARQFAKLVITFGFMGFIIEQYINPIVRNSKHLKGLDLYA - - - - -IERVILKSVNLYVWL 367
  - - - - -
Db 280 FLRLRILEMLFTQVLQGLIQOMVPTIOMSKPKF-DMDYSRIERLLKLVNHLIWL 338

QY 368 CMFYCFHLNLIALLCFDREYKDWNAKSVGDYWRWNNPVHKWVRHIYFPCLR 427
  - - - - -
Db 339 IFYVWLFHSCILNAELMQLFGDREYKDWNSSEVTYFWQNNIPVHKWCIHFYFPMRLR 398

QY 428 SKIPKTLAIITAFVSAVFHELCTAVPCRLKFWAFLGIMFQVPLV-FITVLOEFGST 486
  - - - - -
Db 399 RGSRRMARIGVFLASAFFHELYVSPRLMFLWAFMGMAQIPLANEVGRFFQGNV - - - 455

QY 487 VGNMIFWIFCIGQPCMCVLLYYHD 511
  - - - - -
Db 456 -GNAAVW-LTLIIGQPIAVLMYVHD 478

RESULT 3
DGT1_MOUSE
ID DGT1_MOUSE STANDARD; PRT; 498 AA.
AC Q9Z2A7; Q9D7Q5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Diacylglycerol O-acyltransferase 1 (EC 2.3.1.20) (Diglyceride

```

GenCore version 5.1.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 31, 2003, 00:46:55 ; Search time 38 Seconds
(without alignments)
643.523 Million cell updates/sec

Title: US-09-623-514A-2

Perfect score: 2771

Sequence: 1 MAILDSAGVTTWNGGEF.....QPMCVLLYYHDLNMRKGSMS 520

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	791.5	28.6	488	1	DGTL_HUMAN
2	772.5	27.9	491	1	DGTL_CERAE
3	771.5	27.8	498	1	DGTL_MOUSE
4	762.5	27.5	498	1	DGTL_RAT
5	415.5	15.0	540	1	SOA1_MOUSE
6	412.5	14.9	546	1	SOA1_CRIGR
7	406	14.7	550	1	SOA1_MACFA
8	402	14.5	550	1	SOA1_CERAE
9	397	14.3	550	1	SOA1_HUMAN
10	391.5	14.1	545	1	SOA1_RAT
11	378.5	13.7	642	1	ARE2_YEAST
12	375	13.5	525	1	SOA2_MOUSE
13	357.5	12.9	522	1	SOA2_HUMAN
14	350.5	12.6	526	1	SOA2_CERAE
15	301	10.9	610	1	ARE1_YEAST
16	277.5	10.0	537	1	AREH_SCHPO
17	135	4.9	588	1	YG14_YEAST
18	129	4.7	560	1	NU5M_NEUCR
19	122	4.4	715	1	KN2_HUMAN
20	120	4.3	579	1	TRK1_YEAST
21	120	4.3	1235	1	NU5M_PODAN
22	115	4.2	652	1	DLTB_BACSU
23	113	4.1	395	1	TRK1_SACBA
24	112	4.0	1241	1	HBX3_MOUSE
25	110	4.0	433	1	SRN4_HUMAN
26	109.5	4.0	388	1	YKR4_YEAST
27	109	3.9	618	1	Y147_HAEIN
28	107	3.9	633	1	KN2_MOUSE
29	106.5	3.8	574	1	TRP5_HUMAN
30	106	3.8	973	1	TRP5_RABIT
31	106	3.8	974	1	TRP5_MOUSE
32	106	3.8	975	1	COAA_BPPF1
33	105.5	3.8	437	1	COAA_BPPF1

RESULT 1

ID	DGTL_HUMAN	STANDARD;	PRT;	488 AA.
AC	O75907;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-FEB-2003 (Rel. 41, Last annotation update)			
DE	Diacylglycerol O-acyltransferase 1 (EC 2.3.1.20) (Diglyceride acyltransferase) (ACAT related gene product 1).			
GN	DGAT1 OR DGAT OR AGRP1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98434592; PubMed=9756920;			
RA	Oelkers P., Behari A., Cromley D., Billheimer J.T., Sturley S.L.;			
RT	"Characterization of two human genes encoding acyl coenzyme A:cholesterol acyltransferase-related enzymes.";			
RL	J. Biol. Chem. 273:26765-26771(1998).			
CC	-!- FUNCTION: Catalyzes the terminal and only committed step in triacylglycerol synthesis by using diacylglycerol and fatty acyl CoA as substrates.			
CC	-!- CATALYTIC ACTIVITY: Acyl-CoA + 1,2-diacylglycerol -> CoA + triacylglycerol.			
CC	-!- PATHWAY: Central role in the metabolism of cellular diacylglycerol lipids.			
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic reticulum (By similarity).			
CC	-!- SIMILARITY: BELONGS TO THE STEROL O-ACYLTRANSFERASE FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; AF059202; AAC63997.1; -			
DR	Genew; HGNC:2843; DGAT1.			
DR	MIM; 604900; -			
DR	GO; GO:0008415; F:acyltransferase activity; TAS.			
DR	GO; GO:0004144; F:diacylglycerol O-acyltransferase activity; TAS.			
DR	GO; GO:0006641; P:triacylglycerol metabolism; TAS.			
DR	Pfam; PF03062; MBOAT; 1.			
KW	Transferase; Acyltransferase; Transmembrane; Endoplasmic reticulum.			
FT	TRANSMEM 104..124			POTENTIAL.
FT	TRANSMEM 130..150			POTENTIAL.
FT	TRANSMEM 166..186			POTENTIAL.
FT	TRANSMEM 189..209			POTENTIAL.
FT	TRANSMEM 282..302			POTENTIAL.
FT	TRANSMEM 332..352			POTENTIAL.
FT	TRANSMEM 406..426			POTENTIAL.
FT	TRANSMEM 428..448			POTENTIAL.

ALIGNMENTS

	34	105.5	3.8	580	1	KN2_RAT	P70604	rattus norv
	35	104.5	3.8	431	1	HBX3_HUMAN	P14651	homo sapien
	36	104.5	3.8	445	1	YD26_SCHPO	Q10254	schizosacch
	37	103	3.7	694	1	FR22_DROME	O9vxx3	drosophila
	38	102.5	3.7	367	1	BET3_MESAU	O09029	mesocricetu
	39	101	3.6	525	1	COX1_ASCSU	P24881	ascaris suu
	40	100.5	3.6	637	1	YHE7_YEAST	P46094	homo sapien
	41	100	3.6	333	1	CXCL_HUMAN	P46094	homo sapien
	42	99.5	3.6	318	1	ATH4_ARATH	P92953	arabidopsis
	43	99.5	3.6	577	1	ALG8_YEAST	P40351	saccharomyc
	44	98	3.5	320	1	OME2_HUMAN	Q94255	homo sapien
	45	98	3.5	536	1	MVIN_CHLTR	Q46378	chlamydia t

**THIS PAGE BLANK (USPTO)
BEST AVAILABLE COPY**

THIS PAGE BLANK (USPTO)

DB 208 FYRENVPRFLNNAKSSSVPIPIVNOYL-----YFELAPLILYRDSIPKPIPTVWAGTV 261
 QY 312 ARQFAKLVITFGMGEITIQYINPIVRNSK-HPKGDLLAIAERVLKLSVPMIYWLKMF 370
 DB 262 AMQFAQVQGLLEFYVYIFERLCAPIFRNPKQEPFSARVLV-----LCIF 305

RESULT 14

565208

probable membrane protein YPL189w - Yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein p201

C:Species: Saccharomyces cerevisiae

C>Date: 10-Dec-1994 #sequence_revision 31-May-1996 #text_change 19-Apr-2002

C:Accession: S65208; S65201

R:Reger, M.; Mueller-Auer, S.; Schaefer, M.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S65202

A:Accession: S65208

A:Molecule type: DNA

A:Residues: 1-609 <RIE>

A:Cross-references: EMBL:Z73545; NID:q1370394; PID:e246916; PID:q1370395; MIPS:YPL189w

A:Experimental source: strain S288C (AB972)

R:Benes, V.; Rechmann, S.; Nentwich, U.; Voss, H.; Ansoerge, W.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S65183

A:Accession: S65201

A:Molecule type: DNA

A:Residues: 177-609 <BEN>

A:Cross-references: EMBL:Z73545; MIPS:YPL189w

A:Experimental source: strain S288C (AB972)

C:Genetics:

A:Gene: SGD:GUP2

A:Cross-references: SGD:S0006110

A:Map position: 16L

C:Keywords: transmembrane protein

F:79-95/Domain: transmembrane #status predicted <TM1>
 F:136-152/Domain: transmembrane #status predicted <TM2>
 F:164-180/Domain: transmembrane #status predicted <TM3>
 F:201-217/Domain: transmembrane #status predicted <TM4>
 F:328-344/Domain: transmembrane #status predicted <TM5>
 F:376-392/Domain: transmembrane #status predicted <TM6>
 F:406-422/Domain: transmembrane #status predicted <TM7>
 F:498-514/Domain: transmembrane #status predicted <TM8>
 F:534-550/Domain: transmembrane #status predicted <TM9>
 F:578-594/Domain: transmembrane #status predicted <TM10>

Query Match 5.8%; Score 162; DB 2; Length 609;

Best Local Similarity 19.9%; Pred. No. 3.5e-05;

Matches 94; Conservative 94; Mismatches 151; Indels 134; Gaps 27;

QY 123 SDAIFKQSHAGLFNLCVVLIAVNSRLIENLMKRYGLIRDFWSSRLSDWPLFMCC 182
 DB 122 NSDPQYRFRSFLALILILQI---ILKRVFVFSKPKKPFACGL-----VFVGF 173
 QY 183 I-----SLSTFPLA--AFVTEKLVLOKYSSEPVYIFLHIITMTVELVYVYTLRCDASFL 236
 DB 174 MGINSVYKLFTHAIFETFLANSLKRRRLIAFAFISYGFITL-----FINQKMKMLPF 226
 QY 227 SGVTMLMLTLCIYV-----LKLVSVA-----HTSYDIR-- 263
 DB 227 NIIAIIILSPMDQWYGIYPRWDFPFNFLLRLISTSMDFLERMHEQLSKOPSIDYDDKRP 286
 QY 264 ----SLANNA-----DK-----ANPEVSYYLSKSLAY-----EMVAPTL-- 294
 DB 287 EFRKLSGSTLOTIYESGKNVLEEKERLVAEHHIDYNNINIAITTYAPLVLGPIITF 346
 QY 295 ---CYQPSYPSACIRKG---WVARQFAKLVITFGMGEITIQYINPIVR---NSKHPL 344
 DB 347 NDYLOSENKLPISLTKKNGFALKYFSSILIMEITLHYI---YVGAIAIRAKMNDPFL 403
 QY 345 KQDLLAIAERVLKLSVPMIYWLKMFYCFPHMLNLA-----ELLCGDRREFYKDMWN 398

DB 404 Q-----QAMIALFNLI--NYLKLILPWLFRIMAMVDGIDADENMLRCVDN-----N 449
 QY 399 AASVDYWMWMMNPKWVRIHYEPCLSKIPKTLAIIAFLVSAVHELCIAVPCRLF 458
 DB 450 YSTVG-FRWAMHTSEFKWIRIYVPEGGSN-NKILTSFAVSEFVAIMWIDILRV---LF 504
 QY 459 KLMAFLGIMFQVPLVFTN-YIQERGSTVGMNMFIF-CIFGQP--MCVLL 507
 DB 505 --GWLTVLLILGERTYITNCFSRFRS-----WYRFVCGIGAINICMM 548

RESULT 15

B81409

probable transmembrane transport protein Cj0611c [imported] - Campylobacter jejuni (s

C:Species: Campylobacter jejuni

C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002

C:Accession: B81409

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kellay, J.M.; Churcher, C.; Basham, D.; Chli

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Bar

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals

A:Reference number: A81250; MUID:20150912; PMID:10688204

A:Accession: B81409

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-458 <PAR>

A:Cross-references: GB:AL139075; GB:AL111168; NID:q6967817; PIDN:CAB75247.1; PID:q696

A:Experimental source: serotype O2, strain NCRC 1168

C:Genetics:

A:Gene: Cj0611c

Query Match 5.1%; Score 142; DB 2; Length 458;

Best Local Similarity 20.8%; Pred. No. 0.001;

Matches 86; Conservative 71; Mismatches 165; Indels 92; Gaps 20;

QY 136 ENLCVAVLIAVNSRLIENLMKRYGLIRDFWSSRLSDWPLFMCCISLIFPLAFTV 195
 DB 8 FSLIMIAFFAIY-----WTEKNDYRIQNLIL--IFSYIYILINPYFA--- 49
 QY 196 EKLVILOKYSSEPVYIFLHIITMTVELVYVYTLRCDASFL----- 236
 DB 50 --LVLFYI-----TFPIHFALLIFVRRRIYFATC-MAFITLNLCFEYFPGSINGSDVE 101
 QY 237 ---SGVTMLMLTLCIYVMLKLSVAHTSYDIRSLANAADKANPEVSYYLSKSLAYFM-VA 291
 DB 102 IINFGLERLNIDVLPIDISFYTFSS--ITLVVEYQKRRE-----SFLMLATFISFE 154
 QY 292 PVLCTQPSYPSACIRKGVARQF--AKVIFTGPRGF-----IEQYINPIVRNSKHPLK 345
 DB 155 PVLISGPVIRSSFFEQAVQKREFKRNLIILILVFGIVKVLNLYLGIYAKSKIL--- 210
 QY 346 GDLIAIERVLKLSVPMIYWLKMFYCFPHMLNLI---AELLCGDRREFYDMNNAKSV 402
 DB 211 -DPOSYNFQQLSA--IYAVAIQIYCDSSGYDVLCAPALMLGFLPLPNFMPIYAKNU 267
 QY 403 GDYWRMMNPKVHKWVRHYIFPC--LRSKIPKTLA-IIIAFLVSAVHELCIAVPCORLEK 459
 DB 268 KDFMARWHTLSSTFIRDYIYIPLGNRKGIPTVANIILAFILISGMWHTNTLA-----FI 322
 QY 460 LMAFLGIMFQVPLVFTNLOERFG---SYGNNIMFMIFFIFQOPKCVLLXY 509
 DB 323 VM---GLHIGIVFHLITLTSKFSLOKIPALGREFLQFVGF---TWIFFEY 369

Search completed: August 31, 2003, 04:59:56
 Job time : 51 secs

RESULT 9

T41684

Probable sterol O-acetyltransferase 2 - fission yeast (*Schizosaccharomyces pombe*)C:Species: *Schizosaccharomyces pombe*

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Mar-2002

C:Accession: T41684

R:Medler, H.; Duesternhoeft, A.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, September 1999

A:Reference number: T41684

A:Accession: T41684

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-472 <MEM>

A:Cross-references: EMBL:AL117183; PIDN:CA54864.1; GSPDB:GN00068; SPDB:SPCPIE11.05C

A:Experimental source: strain 972h; clone pl pIE11

C:Genetics:

A:Gene: SPDB:SPCPIE11.05C

A:Map position: 3

C:Superfamily: probable membrane protein YCR048w

Query Match 11.9%; Score 329; DB 2; Length 472;

Best Local Similarity 26.8%; Pred. No. 9.8e-19;

Matches 126; Conservative 74; Mismatches 194; Indels 76; Gaps 18;

OY 100 NADATF-----TTPSVPAHRRARESPLSDAITKOSH-----AGLENLGVVL 143
 DB 15 NLEQTFKGVSETSKIDLRRAAYRPLESPTPS--IFRMYORNAVDFTGFVLEWAV 72
 OY 144 IAVNSRLIEMLMKYGW-LIRDFWESSRSLRDMPLFCISLSIFPLAFLVEVLVQK 202
 DB 73 SIMFMSLEFLELGRVYVGTIRKYRQSNLID--LAAADLAMSMFLAEPFOIFPLG 130
 OY 203 YIS-EPVYFLHIIITMTLVLYPYVTLRC--DSAFSLGVTMLMLTLCIVMLKLVSY-- 255
 DB 131 YLFWYGLGVLYSLITL--LFLSHCVLRCLSNMSTHRAHMFILSHVILMLKLSYVNV 187
 OY 256 -AHNSYDRLS-----ANADKANPEVSY-----VLSKSLATYMAAPT 293
 DB 188 NGWYSYCVHSLNKLQSKRTDLDDESSVEEYECNLNHNHTYPENLITPALDELFMPS 247
 OY 294 LCYQSPYRSACIR-----KGVVARQFAKLVIETFGMFIIEQYINPIYRNSKHPLK 345
 DB 248 LCYQLYYPRTAHVRHIIIECALGFGCIFLLVITSDHFMVPLAKAIRITII--EAPD 304
 OY 346 GDLLYALIER---VLKLSVPMLYWLMCFYCFHMLITIAELLCFGDRREYKDWMAKS 401
 DB 305 ASATYFAIRLGHVAFLEMFPEMLSLVFWVIEGVCNFSAITREAFDRNFYDDMWNMT 364
 OY 402 VGDVWRMNMVHKKMVAHIIYFPCIRSKIRPTLAIITLFLYSAYVHEL---CIAVPCLEF 458
 DB 365 WDOFARWVWKKVHFLLRHVVLP-LNSFMSKSLSTFFTFVSSVLAHELMGCITLKLINGY 423
 OY 459 KLMAFLGIMFOVPLVETINYLOERGSTGVNMIFEFICIFGQPCVLLY 508
 DB 424 GLFF---QMTQIPYIIIOKQKFRVRHRLGNIAFWFSI-IIGIALIALY 469

RESULT 10

S19461

Probable membrane protein YCR048w - yeast (*Saccharomyces cerevisiae*)C:Species: *Saccharomyces cerevisiae*

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Jun-2000

C:Accession: S19461; S19762

R:Grivell, L.A.; de Haan, M.; Maat, M.J.

submitted to the Protein Sequence Database, March 1992

A:Reference number: S19412

A:Accession: S19461

A:Molecule type: DNA

A:Residues: 1-610 <GRI>

A:Cross-references: EMBL:X59720; NID:G1907116; PIDN:CAA42296.1; PID:G1907130; MIPS:YCR04

R:Bohlin-Fukuhara, M.; Buhler, J.M.; Dajman-Fornier, B.; Doltra, C.; Francinques-Gallia

submitted to the Protein Sequence Database, March 1992

A:Reference number: S19396

A:Accession: S19762

A:Molecule type: DNA

A:Residues: 1-328 <BOU>

A:Cross-references: EMBL:X59720; MIPS:YCR048w

C:Genetics:

A:Gene: SGD:ARE1

A:Cross-references: SGD:S0000644; MIPS:YCR048w

A:Map position: 3R

C:Superfamily: probable membrane protein YCR048w

C:Keywords: transmembrane protein

F:186-202/Domain: transmembrane #status predicted <TM1>
 F:228-255/Domain: transmembrane #status predicted <TM2>
 F:262-284/Domain: transmembrane #status predicted <TM3>
 F:289-305/Domain: transmembrane #status predicted <TM4>
 F:370-387/Domain: transmembrane #status predicted <TM5>
 F:401-425/Domain: transmembrane #status predicted <TM6>
 F:453-469/Domain: transmembrane #status predicted <TM7>
 F:538-555/Domain: transmembrane #status predicted <TM8>
 F:590-606/Domain: transmembrane #status predicted <TM9>

Query Match 10.9%; Score 301; DB 2; Length 610;

Best Local Similarity 24.8%; Pred. No. 2.3e-16;

Matches 105; Conservative 76; Mismatches 161; Indels 82; Gaps 17;

OY 156 MKYGL-IR--TDFWSSRSLRDMPLFCISLSIFPLAA--FTVERLYQKYSERPV 209
 DB 195 MELGWAIRCTDYVAYSASA--WN-----KLEIYQVMTDPLFIAMLDLAMELCFPEFV 246
 OY 210 IFHIIII-----TMTLVLYVYVTLKDCSAFLSGVTLMLTLCIYW 249
 DB 247 VFEHVLVKKRIIMKMTGFVAVSIFELAFIPTEPIYV-YFDFENWVRIFLFLHSVVFV 305
 OY 250 LKIVSYA-----HTSYDIRSLANAADKANP-----EVSYY----- 279
 DB 306 MKSHFAYNYGIMDKOLEYSSKQLOKYNESLSPERELIQRKSCDELEINQOTDN 365
 OY 280 -----VSLKSLAYFVAVPTLCYQSPYRSACIRKGVVARQFAKLVIETFGMFIIEQY 333
 DB 366 DEPNNISCSNFEMLFPLVLYQIINYPRTSRIWRVYLEKVC-ALIGTFIMMVTAAQPFM 424
 OY 334 NPI-----VRNSKHPKGLDLVAIERVLT---SVNLLV-VLCMFYCFHMLITIAELL 385
 DB 425 HPVAMKCIQHNTPYEGGWIPYTOEWELHLDMDITGLVTLVTLFTYMLDALLNCAVELT 484
 OY 386 CFGDRREYKDWMAKSVDYWRMNMVHKKMVAHIIYFPCIRSKIRPTLAIITLFLYSAYV 444
 DB 485 RFADRYEFGDMWNCVSEFEFSIRNVVPHKFLLRHVVHSSGALHLSQNTLFTFFISA 544
 OY 445 VFHELCIAPCRLFLMAFLGIMFOVPLVETINYLOERFGSTGVNMIFEFICIFGQPC 504
 DB 545 VFHEKAMFAIRFRVAGYLFMFOLSOFWTALSNTEFLBARPOLSNVVSFGVC-SGPSII 603
 OY 505 VLLY 508
 DB 604 MTLV 607

RESULT 11

T18744

Hypothetical protein B0395.2 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T18744

R:McMurray, A.

submitted to the EMBL Data Library, November 1995

A:Reference number: Z19014

A:Accession: T18744

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-467 <MTL>

A:Cross-references: EMBL:Z68131; PIDN:CAA92217.1; GSPDB:GN00028; CESP:B0395.2

A:Experimental source: clone B0395

RESULT 7

A48026

sterol O-acyltransferase (EC 2.3.1.26) - human

N:Alternate names: ACAT; acyl-coenzyme A cholesterol acyltransferase

C:Species: Homo sapiens (man)

C>Date: 28-May-1999 #sequence_revision 28-May-1999 #text_change 05-May-2000

C/Accession: A59038; A48026

R:Chang, C.C.Y.; Chang, T.Y.

submitted to Genbank, May 1999

A:Description: Molecular cloning and functional expression of human acyl-coenzyme A:cho

A:Reference number: A59038

A:Contents: correction

A:Accession: A59038

A:Molecule type: mRNA

A:Residues: 1-550 <CHAA2>

A:Cross-references: GB:I21934; NID:94878021; PIDN:ACG37532.2; PID:94878022

R:Chang, C.C.Y.; Hub, H.Y.; Gidigan, K.M.; Chang, T.Y.

J. Biol. Chem. 268, 20747-20755, 1993

A:Title: Molecular cloning and functional expression of human acyl-coenzyme A:cholester

A:Reference number: A48026; MUID:94012607; PMID:8407899

A:Accession: A48026

A:Molecule type: mRNA

A:Residues: 1-206, 'R', 208-550 <CHAA1>

A:Cross-references: GB:I21934

C:Genetics:

A:Gene: GDB:SOAT; STAT; ACAT

A:Map position: 1q25-1q25

C:Function:

A:Description: catalyzes the esterification of cholesterol by acyl-CoA

A:Pathway: cholesterol metabolism

A>Note: helps maintain cellular cholesterol homeostasis; plays a role in the development

C:Superfamily: sterol O-acyltransferase

F:Keywords: acyltransferase; cholesterol metabolism; coenzyme A; endoplasmic reticulum;

F:409/491/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.3%; Score 397; DB 1; Length 550;

Best Local Similarity 26.7%; Pred. No. 4e-24;

Matches 144; Conservative 75; Mismatches 177; Indels 144; Gaps 24;

QY 17 GGEFVLDLRKRRKSSDSNGL-----LLSGDNNSBDYGAADVDRIDSYVND 70

DB 70 GSHFDFVTLNLEKSSSLONGGALTTFTLEGEKNHRAKDLRAPE----- 117

QY 71 AGGTANAGDNNCGGNGCGRGGNADATFTYRSPARHARRESPLSSDAIFG 130

DB 118 -GSKIRI-----ARSLDBELLEVDHIRT 141

QY 131 SRAGFNLGVVLIIV-----NSRLIE-NLKKYGM-LIRDFWESSRSLDMPLEMC 182

DB 142 YHMFALLILFLISTLVVDYIDEGRLVLEFSILSYAFGKPTVVM-----TWIMF--- 192

QY 183 ISLSTPLAFTVEKLVQKI-----SEPVYI-----FLHIIITMEVLY-PVYVTLR 230

DB 193 -----LSTFSPYTLFQHWATGYSSHPILRSIFHGFLFMIFQIVGFGPTVYL- 244

QY 231 CDSAPLGVTLMLTICVILKIVSAHT---SYDIRSLANAADKAN-----PEVSYYVSLK 283

DB 245 ---AYTLPPASRIILIFEDIRFPMKASHVRENVPPVLSAKKSSVPIPIYNOQL--- 288

QY 284 SLAYENAVPTLCYQSPYSACIRKGVARQFAKLVIFGFMGFIIEQYINPIVRSK-H 342

DB 299 ---YELFAPTLTYRDSYPRNPVFMGVAMKFAQVGCFFYYVYIERICAPLFRNIKE 355

QY 343 PLKGDLLVIERLKLSV-----PNLYVLMCMFYCFHMLNLALCLCGDREFFKDW 397

DB 356 PPSA-----RLVLICVENSILIPGLILFLFFFAFLHGMNAPMLRSGDMFFKDW 408

QY 398 NAKSVGVYRMNMMPVHKMMVVRHIV--PCLRSKIKPKTALIIIAFLVSAVFHELCTAVGC 455

DB 409 NSTSYSNYIRIVNVVYVHMLIYYAYADFLMFPSKRKSAAMLAIVAVSAVHEIYALAV-C 467

QY 456 R-----LFLKMAFLGIMFQVPLVFTNTYLOERFGSTVGMMIRW-----PFCIFGQ 501

DB 468 LSFEPVPLVFLFMFPGMAFN-----FLVNDKRK---PIVMVLMWTLISFLNGVILCFYSQ 520

RESULT 8

S63350

probable membrane protein YNR019w - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein N3206

C:Species: Saccharomyces cerevisiae

C>Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 20-Jun-2000

C/Accession: S63350

R:Polh, T.M.

submitted to the Protein Sequence Database, April 1996

A:Reference number: S63346

A:Accession: S63350

A:Molecule type: DNA

A:Residues: 1-642 <POH>

A:Cross-references: EMBL:271634; NID:q1302503; PIDN:CAA96298.1; PID:q1302504; MIPS:YN

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:ARE2

A:Cross-references: SGD:S0005302; MIPS:YNR019w

A:Map position: 14R

C:Superfamily: probable membrane protein YCR048w

C:Keywords: transmembrane protein

F:219-235/Domain: transmembrane #status predicted <TM1>

F:295-331/Domain: transmembrane #status predicted <TM2>

F:321-337/Domain: transmembrane #status predicted <TM3>

F:435-451/Domain: transmembrane #status predicted <TM4>

F:483-499/Domain: transmembrane #status predicted <TM5>

F:622-638/Domain: transmembrane #status predicted <TM6>

Query Match 13.7%; Score 378.5; DB 2; Length 642;

Best Local Similarity 25.2%; Pred. No. 1.5e-22;

Matches 133; Conservative 83; Mismatches 176; Indels 135; Gaps 19;

QY 102 DATEFTRPSV-----PAIRRRARE-----SPLSSDAIP 128

DB 125 DMFEHRPSTLSDSVNEPFKTRFVGPPLKEIRRRKELMAMKLNHRRSSPDVDSVG 184

QY 129 KOSHAGFLNLCVV-----VLIAVNSRLIEN--LMKGYL-----IRDFWFS-SRS 172

DB 185 KNDGAAPTPYPTAETSEVTVETTLISNFSGLYVAFWMAIFGAKVALIDYYOHNGS 244

QY 173 LRDMPLFMCCISLIPPLAFTVEKLVQKVISEPVYIFH-----IIT 217

DB 245 FKDSER-----LKMFTNLFTVAVSDMLNTLTSTYFVVGIOYCKMGVLKMGTTGMITFS 298

QY 218 MPEVLYPVYVTLRCDNA---PLSGVTLMLTICVILKIVSA-----HTSYDI 262

DB 299 IYEFLEVFIVMYLTENILKLMHLSKIFLPHSLVLMKMHSPAFNGYILMGIEKELQFSK 358

QY 263 RSLANAADKAN-PEV-----SYVLSLAVFNAVPL 294

DB 359 SALAKYKDSINDKRVGALKESCEPFSLSOSLSDOQKFPNNISAKSFETFTPTL 418

QY 295 CYQSPYSACIRKGVARQFAKLVIFGFMGFIIEQYINPI-----VRNSKHPLKGD 347

DB 419 IYDIEYPRKTEIKWSVLEKIC-AIRGTFILMMIDQILMTVPAMRALAVRNS-----E 471

QY 348 LLYAIEIRVLT-----SVPNLYV-NLCMEFYCFHMLNLALCLCGDREFFKDW 400

DB 472 WSGIDRLRLKMWGLADVDPGFIWMTIDBYLMDALINCVALTFRGDRYFGDMWNCV 531

QY 401 SVGDYRMNMMPVHKMMVVRHIVPCLRS-KIPRTALIIIAFLVSAVFHELCTAVPCRLK 459

DB 532 SWADFSRINIWIPIVHKLHLHYVHSSMSSEFLNKSQATLMTFFLSVYVHELAIVYIKRL 591

QY 460 IMAFLGIMFOVPLVFTNTYLOERFGSTVGMMIRWIFCIGFQPCVCL 506

DB 592 FYLFFQMDPLVPLVLTNTRKMRNRTIIGVIFWALICMGPSVWCL 638

Best Local Similarity 31.5%; Pred. No. 1,6e-50;
Matches 166; Conservative 98; Mismatches 166; Indels 97; Gaps 14;

QY 27 RRRKSRSDSNGLLSGSDNNSPSDVGAPADVRIDSVNDAGTANLAGDNGGSD 86
DB 7 RRRQPSERTSNGSLSSR-----SSFAQ 30
QY 87 NNGGGRGEGCGGNADATFTYPSVAHRARESPSSDAIKKSHA-----GLENLQV 141
DB 31 NNSSRKSEEMRGPEKV-----VH-----TAQDSLSTSGMTNFRGFNLST 75
QY 142 VLIANSRLITENLKKYGLITFDPESS-----SLRNPPLFMCISISTIFPLAFY 196
DB 76 LVLVLSNGRVALENKIVGLITPLQWISTFEVHHYSIMWPMLALICNIDILSPGKE 135
QY 197 KLVLOKTYSEPVVILHIIITTEVLVYVTLACD-SAFLSGVTLLMTCIYMLKIV 255
DB 136 KILRGWLGNGFAAFYISLVLAHLIPVYVTLTKMKRPLMSVVMGYVLEALKETISY 195
QY 256 AHTSY---DIR-----SLANADKA-----NPEVSY-----VSLKSLA 286
DB 196 GHANWWARDARRKITELKTQVTDLAKKTCDFKQFMDLKDLSMHQMAAQYPAKLTLSNIY 255
QY 287 YVMAVPTLCYQPSYRSACIRKGVAROPAKIVITGFGFTIEQYINPYANSKHPKLG 346
DB 256 YMAAPPTLCYEFKFPRLRIRHFLIKRTVELIFLSFLAALVQWVPTVNRSMKPL-S 314
QY 347 DLLVA--TERVVLKESVNLVYVLCMFYCFHMLNLIALCFGEFEYKDMWNAKSVGD 404
DB 315 EMEYRCLERLKLALPHNLITLWLFYTFHFSLMLIALFLADREFRDMNAETTCY 374
QY 405 YVRMNMVYHKMVRHNYFPCLRSKIPTLAIITIAFLVSAVHELCIAVPCFLKMAFL 464
DB 375 FFKSNMIPVHRVRAVHISPMARNNFSKSAFEVFEVSAPFHEYLVSPLKIFRLMSY 434
QY 445 GTFQVPLVETINYLQERGSYVGMIFEFICIGOPCVLLYHD 511
DB 435 GMMGPIPLSTIDKAYR--GGRGTNIYV--LSLIVGOLPLATIMYGD 478

RESULT 5

T01293
hypothetical protein F27F23.25 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 30-Apr-1999
C:Accession: T01293
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, May 1998
A:Description: Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence.
A:Reference number: Z14177
A:Accession: T01293
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-131 <ROD>
A:Cross-references: EMBL:AC003058; NID:g3135250; PID:g3135275
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 2
A:Note: F27F23.25

Query Match 24.2%; Score 670; DB 2; Length 131;
Best Local Similarity 100.0%; Pred. No. 8.6e-47;
Matches 130; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAILDSAGYTVTEGGEFVDLRLRRKRSDDSNGLLGSNNSPSDVGAPADVR 60
DB 1 MAILDSAGYTVTEGGEFVDLRLRRKRSDDSNGLLGSNNSPSDVGAPADVR 60
QY 61 DRIDSVNDAGTANLAGDNGGSDNNGGGRGEGCGGNADATFTYPSVAHRARESP 120
DB 61 DRIDSVNDAGTANLAGDNGGSDNNGGGRGEGCGGNADATFTYPSVAHRARESP 120
QY 121 PLSSDAIFRQ 130

DB 121 PLSSDAIFRQ 130

RESULT 6

sterol O-acyltransferase (EC 2.3.1.26) - mouse
N:Alternate names: acyl-coenzyme A cholesterol acyltransferase
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 02-Jun-2000
C:Accession: I49454; J04617
R:Reinen, P.J.; Oka, K.; Sullivan, M.C.; Chang, T.
J. Biol. Chem. 270, 26192-26201, 1995
A:Title: Molecular cloning of mouse ACAT.
A:Reference number: I49454; M0ID:96064687; PMID:7592824
A:Accession: I49454
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-540 <RES>
A:Cross-references: GB:I42293; NID:g1066809; PID:NAC42075.1; PID:g1066810
R:Green, S.; Steinberg, D.; Quehenberger, O.
Biochem. Biophys. Res. Commun. 218, 924-929, 1996
A:Title: Cloning and expression in Xenopus oocytes of a mouse homologue of the human
A:Reference number: J04617; M0ID:96158386; PMID:8579615
A:Accession: J04617
A:Molecule type: mRNA
A:Residues: 1-194, R', 196-540 <GRE>
A:Cross-references: GB:S81092; NID:g1478335; PID:NAB36050.1; PID:g1478336
A:Experimental source: peritoneal macrophages
C:Comment: This enzyme helps maintain cellular cholesterol homeostasis by catalyzing
y cholesterol and oxygenated sterols.
C:Genetics:
A:Gene: ACAT
A:Map position: 1
C:Superfamily: sterol O-acyltransferase
C:Keywords: acyltransferase; cholesterol; coenzyme A; endoplasmic reticulum; transmem

Query Match 15.0%; Score 415.5; DB 1; Length 540;
Best Local Similarity 28.5%; Pred. No. 1.3e-25;
Matches 135; Conservative 76; Mismatches 149; Indels 113; Gaps 24;

QY 99 GNADATFTYRSPVPAHRA-----RESPSSDAIKKSHA----- 134
DB 81 GALTFTSLIEEMKKNHRRKDLRAPDQKIFISROSL--DELFVDHIRTIVHMFAL 138
QY 135 --LFNLCVAVLAVN--SRLIE-NLMKYGM--LIRDFEMSSSLRDMPLFMCISLSTPP 189
DB 139 LILFVLSTIVVYIDBGRVLEPNLILAVAFGRFPYIV-----TWMAFLSTLSTPPF- 191
QY 190 LAAFVVERLVOKYI-----SEPVVIFL-HIITTEVL-----YPVVTLRCDSAFLS 237
DB 192 -----LFQPMHAGYSKSSHPLIYSLVHGLLFLPVLGVLGFPTVVL---AYTL 238
QY 238 GVTMLILCTIVWLKIVSAHT--SYDIRSLANAADKAN-----PEVSYVSLKSLAYEV 290
DB 239 PPASREILILQIRILMKHSEVRREIPVLAANAEEKSKDPLPVNOYL-----YFLF 292
QY 291 APTLCYQPSYRSACIRKGVAROPAKIVITGFGFTIEQYINPYANSK-HPLKGDLL 349
DB 293 APTLIRDRNYPTPTVYRKYVAMQFLQVGCIFYYIYIFERLCAPLFRNIKQEPISA--- 349
QY 350 VALIERVLKLSV-----PNLYVWLCMFYCFHMLNLIALCFGDRFYMNAKSVGD 404
DB 350 -----RVLVLCVENSLTLPGLILFLESPFAFLHCLMNAFEMFLGDRBMFKDMWNSTSYSN 405
QY 405 YVRMNMVYHKMVRHNYFPCL--RSKIPKTAIIIAFLVSAVHELCIAVPCR----- 456
DB 406 YRTNWNVVDWLVYVYVKDLMLMFEKRSKRSAAVLAVFALSVAHEYALAI-CLSYFYPV 464
QY 457 LFKLNAFLGIMQVPLVETINYLQERGSYVGMIFEFICIGOPCVLLYHD 501
DB 465 LFLVLFMFGMAFN-----FLVNDSKR--PIMINWMAISLPLGLYLICLCTFSQ 510

OY 421 IYFPCRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKIMAFGLIMFOVPLFITNYLQ 480
 |||||||
 DB 421 IYFPCRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKIMAFGLIMFOVPLFITNYLQ 480
 |||||||
 OY 481 ERGSGTVGNMIFEFICIFGQPCVLLYHDLNRRGSM 520
 |||||||
 DB 481 ERGSGTVGNMIFEFICIFGQPCVLLYHDLNRRGSM 520
 |||||||

RESULT 2

H84576
 diacylglycerol O-acyltransferase [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: H84576
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shree, T.P.; Bentley, M.T.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanNieuwen, S.E.; Umayam, L.; Tallon, L.;
 D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: H84576
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-441 <STO>
 A:Cross-references: GB:AE002093; NID:94191775; PIDN:AA01044.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g19450
 A:Map position: 2

Query Match 81.4%; Score 2255.5; DB 2; Length 441;
 Best Local Similarity 83.5%; Pred. No. 1.1e-173;
 Matches 434; Conservative 5; Mismatches 2; Indels 79; Gaps 3;

OY 1 MAIIDSAGVTYTGNGGFEVDLRLRRRSRSDSSNGILSGSDNNSPDVGAADYR 60
 |||||||
 DB 1 MAIIDSAGVTYTGNGGFEVDLRLRRRSRSDSSNGILSGSDNNSPDVGAADYR 60
 |||||||
 OY 61 DRIDSVDNDAGCTANLADGNNGGGNGGRCGCGCAGNADFTYRSPVARRAR 120
 |||||||
 DB 61 DRIDSVDNDAGCTANLADGNNGGGNGGRCGCGCAGNADFTYRSPVARRAR 120
 |||||||
 OY 121 PLSSDAIFQSHAGLNCVVVLIIVNSRLIENLMTKGLRTDFWSSRLRDPLEW 180
 |||||||
 DB 121 PLSSDAIFQSHAGLNCVVVLIIVNSRLIENLMTKGLRTDFWSSRLRDPLEW 180
 |||||||
 OY 181 CCISLIFPLAFTVEKVLQKISRPVIFLIIITMTVEVLPYVTLRDSAFISGV 240
 |||||||
 DB 181 CCISLIFPLAFTVEKVLQKISRPVIFLIIITMTVEVLPYVTLRDSAFISGV 240
 |||||||
 OY 241 LMLTCTIVLKLVSVAHSTYDIRSLANADKANPEVSYVSLKSLAYFWAPLICYOPSY 300
 |||||||
 DB 179 LMLTCTIVLKLVSVAHSTYDIRSLANADKANPEVSYVSLKSLAYFWAPLICYOPSY 300
 |||||||
 OY 301 PPSACIRKQVAVARQFAKLVITFGFMGFIIEOYINPIVRNSKHPKGLDLYAIERVKL 360
 |||||||
 DB 239 PPSACIRKQVAVARQFAKLVITFGFMGFIIEOYINPIVRNSKHPKGLDLYAIERVKL 360
 |||||||
 OY 361 PNLVYWLKCFEYCFHMLNIIAELLCFGRREFKDMNKSVDYDVRMMNMPVHKMVRH 420
 |||||||
 DB 299 PNLVYWLKCFEYCFHMLNIIAELLCFGRREFKDMNKSVDYDVRMMNMPVHKMVRH 420
 |||||||
 OY 421 IYFPCRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKIMAFGLIMFOVPLFITNYLQ 480
 |||||||
 DB 359 IYFPCRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKIMAFGLIMFOVPLFITNYLQ 480
 |||||||
 OY 481 ERGSGTVGNMIFEFICIFGQPCVLLYHDLNRRGSM 520
 |||||||
 DB 408 -----VGNMIFEFICIFGQPCVLLYHDLNRRGSM 520
 |||||||

RESULT 3

T01294
 sterol O-acyltransferase homolog F27F23.26 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 30-Apr-1999
 C:Accession: T01294
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K
 submitted to the EMBL Data Library, May 1998
 A:Description: Arabidopsis thaliana chromosome II BAC F27F23 genomic sequence.
 A:Reference number: Z14177
 A:Accession: T01294
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-231 <ROU>
 A:Cross-references: EMBL:AC003058; NID:93135250; PID:93135276
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 2
 A:Introns: 44/3; 70/3; 104/3; 150/2; 177/3; 184/3; 205/3
 A:Note: F27F23.26

Query Match 43.4%; Score 1202; DB 2; Length 231;
 Best Local Similarity 97.4%; Pred. No. 2.7e-89;
 Matches 222; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

OY 230 RCDSAFLSGVTLMLTLCIYWLKLVSAHSTYDIRSLANADKANPEVSYVSLKSLAYFW 289
 |||||||
 DB 3 RCDSAFLSGVTLMLTLCIYWLKLVSAHSTYDIRSLANADKANPEVSYVSLKSLAYFW 289
 |||||||
 OY 290 VAPTLCTOPSPYPSACIRKQVAVARQFAKLVITFGFMGFIIEOYINPIVRNSKHPKGLD 349
 |||||||
 DB 63 VAPTLCTOPSPYPSACIRKQVAVARQFAKLVITFGFMGFIIEOYINPIVRNSKHPKGLD 349
 |||||||
 OY 350 YAIERVLKLSVPMIYWLKCFEYCFHMLNIIAELLCFGRREFKDMNKSVDYDVRMM 409
 |||||||
 DB 123 YAIERVLKLSVPMIYWLKCFEYCFHMLNIIAELLCFGRREFKDMNKSVDYDVRMM 409
 |||||||
 OY 410 NMPVHKMVRHIIYFPCRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKIMAFGLIM 455
 |||||||
 DB 183 NMPVHKMVRHIIYFPCRSKIPKTLAIIIAFLVSAVHELCIAVPCRLFKIMAFGLIM 455
 |||||||

RESULT 4

T19027
 probable sterol O-acyltransferase (EC 2.3.1.26) H19N07.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 05-May-2000
 C:Accession: T19027; T23106
 R:Kershaw, J.
 submitted to the EMBL Data Library, June 1996
 A:Reference number: Z19061
 A:Accession: T19027
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-496 <WIL>
 A:Cross-references: EMBL:Z75526; PIDN:CAA9773.1; GSPDB:GN00023; CESP:H19N07.4
 A:Experimental source: clone C06H2
 R:Dobson, R.
 submitted to the EMBL Data Library, March 1997
 A:Reference number: Z19678
 A:Accession: T23106
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-496 <WIL>
 A:Cross-references: EMBL:Z92835; PIDN:CAB07399.1; GSPDB:GN00023; CESP:H19N07.4
 A:Experimental source: clone H19N07
 C:Genetics:
 A:Gene: CESP:H19N07.4
 A:Map position: 5
 A:Introns: 5/3; 43/2; 77/3; 139/1; 171/3; 456/1
 A:Superfamily: sterol O-acyltransferase
 C:Keywords: acyltransferase; coenzyme A

Query Match 26.1%; Score 724.5; DB 2; Length 496;

P. N. S. S. S.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: August 31, 2003, 04:34:04 ; Search time 46 Seconds

(without alignments)
1087.124 Million cell updates/sec

Title: US-09-623-514a-2

Perfect score: 2771

Sequence: 1 MAILDSAGVTVTENGGEF.....QPMCVLLYHDLNRRKGSMS 520

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2771	100.0	520	2	diacylglycerol O-a
2	2255.5	81.4	441	2	diacylglycerol O-a
3	1202	43.4	231	2	sterol O-acyltrams
4	724.5	26.1	496	2	probable sterol O-
5	670	24.2	131	2	hypothetical prote
6	415.5	15.0	540	1	sterol O-acyltrams
7	397	14.3	550	1	sterol O-acyltrams
8	378.5	13.7	472	2	probable membrane
9	329	11.9	472	2	probable sterol o-
10	301	10.9	610	2	probable membrane
11	279	10.1	467	2	hypothetical prote
12	277.5	10.0	537	2	probable sterol O-
13	165.5	6.0	305	2	sterol O-acyltrams
14	162	5.8	609	2	probable membrane
15	142	5.1	458	2	probable membrane
16	139.5	5.0	404	2	probable membrane
17	134.5	4.9	865	2	probable membrane
18	129	4.7	560	2	probable membrane
19	126	4.5	495	2	probable membrane
20	122	4.4	715	2	probable membrane
21	121.5	4.4	450	2	conserved hypothet
22	121.5	4.4	450	2	conserved hypothet
23	121.5	4.4	1170	2	probable 2-acylgly
24	120	4.3	1235	1	probable transpor
25	118.5	4.3	865	2	probable membrane
26	117.5	4.2	499	2	alginat o-acetyl
27	116	4.2	1689	2	sodium channel pro
28	115	4.2	652	2	NAH2 dehydrogenas
29	113	4.1	395	2	diLB protein - Bac

30	112	4.0	352	2	T48903	wax synthase (limp
31	112	4.0	1241	2	JU0466	potassium transpor
32	111	4.0	395	2	B96610	hypothetical prote
33	111	4.0	420	2	D83556	probable coat prot
34	109.5	4.0	263	2	AE1557	hypothetical prote
35	109.5	4.0	388	2	JN0605	somatostatin recep
36	109	3.9	618	1	S38004	probable transpor
37	108.5	3.9	783	2	A96825	hypothetical prote
38	108	3.9	527	2	G64626	alginat O-acetyl
39	107	3.9	334	2	T19955	hypothetical prote
40	107	3.9	395	2	AE2349	hypothetical prote
41	107	3.9	633	2	T64143	hypothetical prote
42	106.5	3.8	1553	2	T18502	hypothetical prote
43	106	3.8	546	2	S52053	cytochrome-c oxida
44	106	3.8	2895	2	H85362	hypothetical prote
45	105.5	3.8	437	2	S15144	hypothetical prote

ALIGNMENTS

used this for R (ist)
as reference
500

RESULT 1	
T52584	diacylglycerol O-acyltransferase (EC 2.3.1.20) [validated] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)	
C:Date: 24-Oct-2000 #sequence_revision 24-Oct-2000 #text_change 17-Nov-2000	
C:Accession: T52584	
R:Hills, M.J.; Lu, C.; Hobbs, D.H.	
FEBS Lett. 452, 145-149, 1999	
A:Title: Cloning of a cDNA encoding diacylglycerol acyltransferase from Arabidopsis t	
A:Reference number: 226127	
A:Accession: T52584	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-520 <Hills>	
A:Cross-References: EMBL:AJ11831; PIDN:CA644774.1	
A:Experimental source: cultivar Columbia	
A:Function:	
A:Description: EC 2.3.1.20 [validated, MUID:99313150]; catalyzes the final acylation	
A:Pathway: triacylglycerol biosynthesis	
C:Keywords: acyltransferase; coenzyme A	
Query Match	100.0%; Score 2771; DB 2; Length 520;
Best Local Similarity	100.0%; Pred. No. 4.5e-215;
Matches 520; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 MAILDSAGVTVTENGGEFVDLRLRRKSSDSSNGILLSGSNNNSPDVGAADVR 60
DB	1 MAILDSAGVTVTENGGEFVDLRLRRKSSDSSNGILLSGSNNNSPDVGAADVR 60
OY	61 DRIDSVDNDAGCTANLADGNNGGDNNGGGEGRGGMADATFTYRSPVAHRRARS 120
DB	61 DRIDSVDNDAGCTANLADGNNGGDNNGGGEGRGGMADATFTYRSPVAHRRARS 120
OY	121 PLSSDAIFKOSHAGLFNLCVVLIVNSRLITENIMTKYGLIRTDFWSSRSRLRMPLE 180
DB	121 PLSSDAIFKOSHAGLFNLCVVLIVNSRLITENIMTKYGLIRTDFWSSRSRLRMPLE 180
OY	181 CCISSISIFPLAFYTEKVLVLOKYYISEPVYIFHHITTEVLYPYVYVLRCSAFLSGVT 240
DB	181 CCISSISIFPLAFYTEKVLVLOKYYISEPVYIFHHITTEVLYPYVYVLRCSAFLSGVT 240
OY	241 LMLLCIYWLKIVSYAHTSYDRLSLANADKANPESSYVSLSLAYFVAATFLCYOPSY 300
DB	241 LMLLCIYWLKIVSYAHTSYDRLSLANADKANPESSYVSLSLAYFVAATFLCYOPSY 300
OY	301 PSACIRKGVARQPAKLVIFTFMGFTIEQYINPIVNRSKHPLGDLLYALERYLKLVS 360
DB	301 PSACIRKGVARQPAKLVIFTFMGFTIEQYINPIVNRSKHPLGDLLYALERYLKLVS 360
OY	361 PMLYVWLCMFYCFEFLHMLITLAEELCFDGRREYKRWMAKSGVDYWRKMMNPHVHWVRH 420
DB	361 PMLYVWLCMFYCFEFLHMLITLAEELCFDGRREYKRWMAKSGVDYWRKMMNPHVHWVRH 420

THIS PAGE BLANK (USPTO)

RESULT 15
CENB HUMAN STANDARD; PRT; 599 AA.
AC P07159;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Major centromere autoantigen B (Centromere protein B) (CENP-B).
GN CENPB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91372020; PubMed=1893793;
RA Sullivan K.F., Glaes C.A.;
RT "CENP-B is a highly conserved mammalian centromere protein with
RT homology to the helix-loop-helix family of proteins.";
RL Chromosoma 100:360-370(1991).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavridis G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Buttrill W.D., Butler A.P., Garder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A.G., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gillman R., Hall R.E.,
RA Hammond S., Harley J.R., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leveleslho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McNulty A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,
RA Rice C.M., Rose M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
RA Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sultson J.E.,
RA Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromens A.C., Vautin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:665-671(2001).
[3]
RP SEQUENCE OF 6-599 FROM N.A.
RX MEDLINE=97166180; PubMed=2435739;
RA Bernshaw W.C., Sullivan K.F., Machlin P.S., Cooke C.A.,
RA Kaiser D.A., Pollard T.D., Rochfield N.F., Cleveland D.W.;
RT "Molecular cloning of cDNA for CENP-B, the major human centromere
RT autoantigen.";
RL J. Cell Biol. 104:817-829(1987).
[4]
RP SUBUNITS, AND DOMAINS.
RX MEDLINE=93107144; PubMed=1469042;
RA Yoda K., Kitagawa K., Masumoto H., Muro Y., Okazaki T.;
RT "A human centromere protein, CENP-B, has a DNA binding domain
RT containing four potential alpha helices at the NH2 terminus, which is
RT separable from dimerizing activity.";
RL J. Cell Biol. 119:1413-1427(1992).
[5]
RP STRUCTURE BY NMR OF 1-56.
RX MEDLINE=98119825; PubMed=9451007;
RA Iwahara J., Kitagawa T., Kitagawa K., Masumoto H., Okazaki T.,
RA Yokoyama S.;

RT "A helix-turn-helix structure unit in human centromere protein B
RT (CENP-B).";
RL EMBL J. 17:827-837(1998).
CC -1- FUNCTION: INTERACTS WITH CENTROMERIC HETEROCHROMATIN IN
CC CHROMOSOMES AND BINDS TO A SPECIFIC SUBSET OF ALPHOID SATELLITE
CC DNA, CALLED THE CENP-B BOX. MAY ORGANIZE ARRAYS OF CENTROMERE
CC SATELLITE DNA INTO A HIGHER ORDER STRUCTURE WHICH THEN DIRECTS
CC CENTROMERE FORMATION AND KINETOCORE ASSEMBLY IN MAMMALIAN
CC CHROMOSOMES.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: Contains 1 CENPB domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X55039; CAA38879.1; -;
DR EMBL: AL109804; CAC17547.1; -;
DR EMBL: X05299; CAA28918.1; -;
DR PIR: S18735; S18735.
DR PDB: 1BW6; 07-OCT-98.
DR PDB: 1HTV; 11-JAN-02.
DR Genew: HGNC:1852; CENPB.
DR MIM: 117140; -;
DR GO: GO:0005698; C:centromere; NAS.
DR GO: GO:0003682; F:chromatin binding activity; NAS.
DR GO: GO:0003696; F:satellite DNA binding activity; NAS.
DR GO: GO:0000069; P:centromere/kinetochore complex maturation; NAS.
DR InterPro: IPR004875; CENPB-B.
DR InterPro: IPR006600; CENPB.
DR Pfam: PF04218; CENP-B_N; 1.
DR Pfam: PF03184; DDB; 1.
DR SMART: SM00674; CENPB; 1.
KM Chromosomal protein; Nuclear protein; DNA-binding; Centromere;
KM 3D-structure.
FT DNA BIND 1 125
FT DOMAIN 404 465 GLU-RICH (ACIDIC).
FT DOMAIN 508 538 ASP/GLU-RICH (ACIDIC).
FT CONFLICT 583 583 R -> M (IN REF. 3).
FT CONFLICT 592 593 VR -> LL (IN REF. 3).
FT HELIX 10 21
FT HELIX 28 35
FT TURN 39 40
FT HELIX 41 47
FT HELIX 49 52
FT TURN 53 53
SQ SEQUENCE 599 AA; 65171 MW; 9B4B7DB957A914AA CRC64;
Query Match 16.0%; Score 138; DB 1; Length 599;
Best local similarity 26.3%; Pred. No. 0.11;
Matches 36; Conservative 18; Mismatches 37; Indels 46; Gaps 3;
QY 38 EEEQQQQLRQGEHEHTARNGEVGVPRPGQNDGQGLENNNRISVDEDS -GNG 96
DB 406 EEEEEEEEEEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGEGE
QY 97 EDEDEDESHAGE-----ODEDEDEDEEM 118
DB 459 EEEDEDESSSEGLEAEEDMACGVVEAGSGFGAYGAGERAOCPTLHFLGGSDSDSDSEED 518
QY 119 DQSSDPFDGSDSDSSRED 135
DB 519 DEEDDEDEDDDDDEED 535
Search completed: December 10, 2003, 19:57:55
Job time : 20 secs

SQ SEQUENCE 765 AA; 89509 MW; 006A20A7F968DB6A CRC64;
 Query Match 16.2%; Score 139.5; DB 1; Length 765;
 Best Local Similarity 22.8%; Pred. No. 0.11;
 Matches 41; Conservative 38; Mismatches 58; Indels 43; Gaps 6;

QY 1 MN0ELLVSGSRRRTGSLRGNPSSQVDEQKMRVVEEEOQ-----QLR 46
 DB 593 LKEMWVIGSRMORI-----SOSQKHYYKLAEBQQRKYKHLDLWVKSLSPODR 642
 QY 47 QOEERTARN---GEVGVGPRPGQNDSCQQLLENNNRFTSVDEDSNGOEDEDE 102
 DB 643 AAYEYISNRKKNMTKLRGPRKPSRTTLQSKSESEDDDE---EEEDDEEEEDDD 698
 QY 103 EH-----AGQDEDEDEE---EEMDQSDDPQSDSRDEDEHHTNSVTNSSLVD 151
 DB 699 ENGDSEEDGDDSSSESEDESDDDDDDDDDDDDDDDDDDDDDDESDNESESSSSSSSGD 758

RESULT 14
 ID CNM4_BOVIN STANDARD; PRT; 1394 AA.
 AC Q28181; Q03861; Q28082;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 240 kDa protein of rod photoreceptor CNM-channel [contains: Glutamic acid-rich protein (GARP); Cyclic-nucleotide-gated cation channel 4 (CNM channel 4) (CNM-4) (Cyclic nucleotide-gated cation channel modulatory subunit)].
 DE CNM3 OR CNM4.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 NC NCB1_TaxID=9913;
 [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=96009859; PubMed=7546742;
 RA Goetzes S., Colville C., Mueller F., Dose A., Godde M., Molday L., Kaupp U.B., Molday R.S.;
 RA "A 240 kDa protein represents the complete beta subunit of the cyclic nucleotide-gated channel from rod photoreceptor";
 RT Neuron 15:627-636 (1995).
 RL [2]
 RP SEQUENCE OF 454-1394 FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=96198098; PubMed=8626431;
 RA Biel M., Zong X., Ludwig A., Sauter A., Hofmann F.;
 RA "Molecular cloning and expression of the modulatory subunit of the cyclic nucleotide-gated cation channel";
 RT J. Biol. Chem. 271:6349-6355 (1996).
 RL [3]
 RP SEQUENCE OF 1-590 FROM N.A.
 RC TISSUE=Retina;
 RA Sugimoto Y., Yatsunami K., Tsujimoto M., Khourana H.G., Ichikawa A.;
 RA Submitted (Xxx-1991) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBUNIT: FORMS FUNCTIONAL HETEROLOGIC CHANNELS WITH CNM3.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC (CNM4D=CNM4C; CNM4E = 20:2:1) in testis;
 CC Name=CNM4C;
 CC IsoId=Q28181-1; Sequence=Displayed;
 CC Name=CNM4D;
 CC IsoId=Q28181-2; Sequence=VSP_001109;
 CC Name=CNM4E;
 CC IsoId=Q28181-3; Sequence=VSP_001108;
 CC -1- TISSUE SPECIFICITY: RETINA, TESTIS, KIDNEY, HEART, AND BRAIN.
 CC -1- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X89626; CA61769.1; -
 CC EMBL; X94707; CA6367.1; -
 CC EMBL; M61185; AAA30536.1; -
 CC PIR; A40437; A40437.
 DR InterPro: IPR000595; CNM binding.
 DR InterPro: IPR005821; Ion trans.
 DR Pfam; PF00027; CNM_binding; 1.
 DR Pfam; PF00520; Ion_trans; 1.
 DR SMART; SMO0100; CNM; 1.
 DR PROSITE; PS00888; CNM_BINDING_1; 1.
 DR PROSITE; PS00889; CNM_BINDING_2; 1.
 DR PROSITE; PS50042; CNM_BINDING_3; 1.
 KW Ionic channel; Ion transport; CAMP-binding; Transmembrane; Multigene family; Alternative splicing.
 KW CHAIN 1 590
 FT CHAIN 454 1394
 FT DOMAIN 1 767
 FT TRANSMEM 768 786
 FT DOMAIN 801 819
 FT TRANSMEM 820 844
 FT DOMAIN 845 864
 FT TRANSMEM 865 901
 FT TRANSMEM 902 924
 FT DOMAIN 925 968
 FT TRANSMEM 969 988
 FT DOMAIN 1073 1093
 FT TRANSMEM 1094 1394
 FT NP BIND 1081 1219
 FT BINDING 1141 1141
 FT BINDING 1153 1153
 FT CARBOHYD 1067 1067
 FT VAAPPLIC 515 532
 FT VAAPPLIC 522 530
 FT CONFLICT 341 341
 FT CONFLICT 454 465
 FT CONFLICT 482 482
 FT CONFLICT 499 499
 FT CONFLICT 572 590
 FT CONFLICT 1283 1283
 FT CONFLICT 1289 1289
 FT CONFLICT 1336 1336
 FT CONFLICT 1338 1338
 SQ SEQUENCE 1394 AA; 155064 MW; EB6DA559BB374AA7 CRC64;
 Query Match 16.1%; Score 138.5; DB 1; Length 1394;
 Best Local Similarity 24.6%; Pred. No. 0.22;
 Matches 42; Conservative 35; Mismatches 43; Indels 51; Gaps 7;

QY 17 GSLKGNPSSQV---DEQMRVVE-----EEQQQLRGQEEHTARNGEV 59
 DB 321 GSTSTSPRTSAAPADEKGVVEQTPELPRIQEKDEBEKEDGEEBEKGEKKEE 380
 QY 60 VG--VEPPGQNDSCQQLLENNNRFTSV-----EDSSNGOEDEE---DEEHAGEQ 108
 DB 381 EGEKEKEGEKKE 440
 QY 109 DEED-----EEEMDQSDP-----DPQSDSSDEDE 136
 DB 441 EEEBGRKEVEGREEDDEEBQDHSVILDSYLVQSEEDRSSESTODQ 491

```

RA Loebl A.D., Seaver L., Bonnefont J.-P., Romano C., Fichera M.,
RA Lacombe D., Hanauer A., Philip N., Schwartz C.E., Fontes M.,
RT "Evaluation of a mutation screening strategy for sporadic cases of
RT ATR-X syndrome.";
RL J. Med. Genet. 36:183-186 (1999).
RN (17)
RN VARIANTS ATR-X SER-179; LEU-190; ILE-194; CYS-246; PHE-1552; SER-1645
RP AND CYS-1847.
RX MEDLINE=20451413; PubMed=10995512;
RA Wada T., Kubota T., Fukushima Y., Satoh S.;
RT "Molecular genetic study of Japanese patients with X-linked alpha-
RT thalassemia/mental retardation syndrome (ATR-X).";
RL Am. J. Med. Genet. 94:242-248 (2000).
CC -1- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN BRAIN
CC DEVELOPMENT AND FACIAL MORPHOGENESIS.
CC -1- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEXIN V IN A CALCIUM AND
CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
CC HEREROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
CC INTERACTING WITH HPL.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Name=1;
CC IsoId=P46100-1; Sequence=Displayed;
CC Name=2;
CC IsoId=P46100-2; Sequence=VSP_000575;
CC Name=3;
CC IsoId=P46100-3; Sequence=VSP_000574;
CC Name=4;
CC IsoId=P46100-4; Sequence=VSP_000576;
CC Name=5;
CC IsoId=P46100-5; Sequence=VSP_000574, VSP_000576;
CC -1- TISSUE SPECIFICITY: Ubiquitous.
CC -1- DISEASE: Defects in ATRX are the cause of X-linked alpha-
CC thalassemia/mental retardation syndrome (ATR-X) [MIM:301040]. ATR-
CC X is an X-linked disorder comprising severe psychomotor
CC retardation, facial dysmorphism, urogenital abnormalities, and
CC alpha-thalassemia. An essential phenotypic trait are hemoglobin H
CC erythrocyte inclusions.
CC -1- DISEASE: Defects in ATRX are the cause of Sutherland-Haan X-linked
CC mental retardation syndrome (SHS) [MIM:309470]. It is
CC characterized by severe mental retardation with spastic
CC paraplegia, microcephaly, short stature and cryptorchidism.
CC -1- DISEASE: Defects in ATRX are a cause of Smith-Fineman-Myers
CC syndrome (SPM) [MIM:309580]. Clinical features include severe
CC mental retardation, microcephaly, growth failure, facial anomalies
CC and bilateral cryptorchidism. Due to the clinical overlap with
CC ATR-X syndrome, some patients originally diagnosed as having SPM,
CC might be affected by a variant of ATR-X syndrome which lack
CC hemoglobin H inclusions.
CC -1- DISEASE: Defects in ATRX are the cause of Carpenter-Waziri
CC syndrome (CWS), an X-linked recessive condition characterized by
CC moderate mental retardation, short stature, brachydactyly with
CC excessive skin creases, and widening of the knuckles.
CC -1- DISEASE: Defects in ATRX are the cause of Juberger-Marsidi syndrome
CC (JM) [MIM:309590]. JM is a rare X-linked recessive disease
CC characterized by severe mental retardation, growth failure,
CC sensorineural deafness, microgenitalism and early death.
CC -1- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -1- SIMILARITY: Contains 1 PHD-type zinc finger.
CC -----
CC
Query Match 16.5%; Score 141.5; DB 1; Length 2492;
Best Local Similarity 27.4%; Pred. No. 0.26;
Matches 49; Conservative 21; Mismatches 42; Indels 67; Gaps 7
CC
12 RRRTGGSLRGNSPSQVDEQGNRV-----VEEEOOOLROQ-----EE 50
Db 1300 KKRGTQKQVENRG-----DEAKNQNSDSDSSESKPRRRHRLRLKLTYSDDSEGE 135
51 EHT--ANRGAVGVVQPRPGGQNDQ-----GQOLEEN 80

```

```

Db      1356 KKTPEKHEKVKGNRRKVSSEDESDDFQSGCVSEVSSSEDEDFQRTSRSAKKAELLEN 1415
Qy      81 -----NNRFISVDESSGNGOEBOEDEHAGODEDEEHEMDQESDDPQSDS 131
Db      1416 QRSYKQKKRRIRIVQDSSESNKSNSENEEEEEKEEEEEEEEEEEEEEE-----DENDS 1470

RESULT 13
UBF1_MOUSE
ID      UBF1_MOUSE      STANDARD;      PRT;      765 AA.
AC      P25976;
DT      01-MAY-1992 (Rel. 22, Created)
DT      01-MAY-1992 (Rel. 22, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      Nucleolar transcription factor 1 (Upstream binding factor 1) (UBF-1).
GN      UBF1 OR TCFUBF OR UBF1 OR UBF-1.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Euteheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI TaxID=10090;
[1]
SEQUENCE FROM N.A. (UBF1 AND UBF2).
STRAIN=C3H/He;
MEDLINE=91367658; PubMed=1891354;
Hlaatske K., Nishimura T., Maeda Y., Hanada K.I., Song C.Z.,
Miyamatsu M.;
"Cloning and structural analysis of cDNA and the gene for mouse
transcription factor UBF.";
Nucleic Acids Res. 19:4631-4637(1991).
-1- FUNCTION: UBF RECOGNIZES THE RIBOSOMAL RNA GENE PROMOTER AND
ACTIVATES TRANSCRIPTION MEDIATED BY RNA POLYMERASE I THROUGH
COOPERATIVE INTERACTIONS WITH THE SPECIES-SPECIFIC FACTOR SL1. IT
BINDS SPECIFICALLY TO THE UPSTREAM CONTROL ELEMENT.
-1- SUBUNIT: Homodimer.
-1- SUBCELLULAR LOCATION: Nuclear.
-1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=UBF1; Synonyms=Long;
IsoId=P25976-1; Sequence=Displayed;
Name=UBF2; Synonyms=Short;
IsoId=P25976-2; Sequence=VSP_002194;
-1- SIMILARITY: Contains 6 HMG box domains.
-----
CC      CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
-----
DR      EMBL; X60831; CAA43222.1; -.
DR      PIR; S22314; S22314.
DR      HSSP; P07155; IHMF.
DR      MGD; MGI:98512; IHMF.
DR      GO; GO:0005730; C:nucleolus; IDA.
DR      GO; GO:0005515; F:protein binding activity; IPI.
DR      InterPro; IPR000910; HMG_12_box.
DR      Pfam; PF00505; HMG_box; 5.
DR      SMART; SM00398; HMG; 6.
DR      PROSITE; PS50118; HMG_BOX_2; 6.
KW      Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW      Repeat; Alternative splicing.
FT      DNA BIND 112 180 HMG BOX 1.
FT      DNA BIND 196 264 HMG BOX 2.
FT      DNA BIND 298 362 HMG BOX 3.
FT      DNA BIND 407 475 HMG BOX 4.
FT      DNA BIND 482 549 HMG BOX 5.
FT      DNA BIND 568 634 HMG BOX 6.
FT      DOMAIN 675 765 HMG BOX 6.
FT      VARSP LIC 221 257 ASP/GLU/SER-RICH (ACIDIC).
                          Missing (in isoform UBF2).
                          /Fld=VSP_002194.

```


DB 643 AAYKEYISNKKRNMKTLRGPNKSGRTTLQSKSSEEDD-----EEDDDDDDEEEEDD 697

OY 103 EH-----AGEDEDEDEEEMDQSDPDQSD--SSRDEHTHNTNNTSSS 148

DB 698 ENGDSSEDEGSSSESSSEDESEDEGDEDDDDDDDEDEDEDEDEDESESSSSS 753

RESULT 12

ATRX_HUMAN

ID ATRX_HUMAN STANDARD; PRT; 2492 AA.

AC P46100; P51068; Q15886; Q9H021; Q9NTS3;

DT 01-NOV-1995 (Rel. 32, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked nuclear protein) (XNP) (Znt-HX).

GN ATRX OR RAD54L OR XH2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

OC NCBI_taxid=9606;

OK

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS 1, 2, 3, 4, AND 5), VARIANT SER-1860, AND

RP VARIANTS ATR-X.

RX MEDLINE=97123494; PubMed=8968741;

RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,

RA Gibbons R.J.;

RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations point to a common mechanism underlying the ATR-X syndrome.";

RT Hum. Mol. Genet. 5:1899-1907(1996).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).

RX MEDLINE=97386582; PubMed=9244431;

RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,

RA Colleaux L., Schwartz C., Fontes M.;

RT "Determination of the genomic structure of the XNP/ATRX gene encoding a potential zinc finger helicase.";

RT Genomics 43:149-155(1997).

RN [3]

RP SEQUENCE OF 860-2492 FROM N.A.

RX MEDLINE=95179111; PubMed=7874112;

RA Stayton C.L., Dabovic B., Giuliano M., Gecz J., Broccoli V.,

RA Giovannazzi S., Bossoiasco M., Monaco L., Raetan S., Boncinelli E.,

RA Bianchi M.E., Consalez G.G.;

RT "Cloning and characterization of a new human Xq13 gene, encoding a putative helicase.";

RT Hum. Mol. Genet. 3:1957-1964(1994).

RN [4]

RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.

RX MEDLINE=94214473; PubMed=8162050;

RA Gecz J., Pollard H., Consalez G., Villard L., Stayton C.L.,

RA Millaesau P., Khrestchatsky M., Fontes M.;

RT "Cloning and expression of the murine homologue of a putative human X-linked nuclear protein gene closely linked to PKI in Xq13.3.";

RT Hum. Mol. Genet. 3:39-44(1994).

RN [5]

RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.

RX MEDLINE=95211835; PubMed=7697714;

RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;

RT "Mutations in a putative global transcriptional regulator cause X-linked mental retardation with alpha-thalassemia (ATR-X syndrome).";

RT Cell 80:837-845(1995).

RN [6]

RP SEQUENCE OF 1375-2492 FROM N.A.

RA Pearce A., Chapman J.;

RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

RN [7]

RP EZH2 BINDING.

RX MEDLINE=98167853; PubMed=9499421;

RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,

RA Colleaux L.;

RT "Specific interaction between the XNP/ATR-X gene product and the SET domain of the human EZH2 protein.";

RN [8]

RP Hum. Mol. Genet. 7:679-684(1998).

RN [9]

RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC

RP HETEROCHROMATIN.

RX MEDLINE=20040663; PubMed=10570185;

RA McEwell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,

RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,

RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;

RT "Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes.";

RT Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).

RN [9]

RP DISEASE.

RX MEDLINE=20231147; PubMed=10751095;

RA Villard L., Fontes M., Ades L.C., Gecz J.;

RT "Identification of a mutation in the XNP/ATR-X gene in a family reported as Smith-Fineman-Myers syndrome.";

RT Am. J. Med. Genet. 91:83-85(2000).

RN [10]

RP VARIANT ATR-X SER-1713.

RX MEDLINE=97196774; PubMed=9043863;

RA Villard L., Lacombe D., Fontes M.;

RT "A point mutation in the XNP gene, associated with an ATR-X phenotype without alpha-thalassemia.";

RT Eur. J. Hum. Genet. 4:316-320(1996).

RN [11]

RP VARIANT JM GIN-2131.

RX MEDLINE=96224392; PubMed=8630485;

RA Villard L., Gecz J., Mattei J.-F., Fontes M., Saugier-Verber P.,

RA Munnich A., Lyonnet S.;

RT "XNP mutation in a large family with Juberger-Marsidi syndrome.";

RT Nat. Genet. 12:359-360(1996).

RN [12]

RP VARIANTS ATR-X.

RX MEDLINE=97467722; PubMed=9326931;

RA Gibbons R.J., Bachoo S., Picketts D.J., Affimos S., Azenbauer B.,

RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,

RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Stanley S.F.,

RA Higgs D.R.;

RT "Mutations in transcriptional regulator ATRX establish the functional significance of a PHD-like domain.";

RT Nat. Genet. 17:146-148(1997).

RN [13]

RP VARIANT ATR-X LEU-246.

RX MEDLINE=20123062; PubMed=10660327;

RA Fichera M., Romano C., Castiglia L., Falla P., Ruberto C., Amata S.,

RA Greco D., Cardoso C., Fontes M., Ragusa A.;

RT "New mutations in XNP/ATR-X gene: a further contribution to genotype/phenotype relationship in ATR/X syndrome.";

RT Hum. Mutat. 12:214-214(1998).

RN [14]

RP VARIANT SHS-LYS-1742.

RX MEDLINE=99347960; PubMed=10417298;

RA Lossi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,

RA Prieto F., Fontes M., Martinez F.;

RT "Mutation of the XNP/ATR-X gene in a family with severe mental retardation, spastic paraplegia and skewed pattern of X inactivation: demonstration that the mutation is involved in the inactivation bias.";

RT Am. J. Hum. Genet. 65:558-562(1999).

RN [15]

RP VARIANT CMS THR-2050.

RX MEDLINE=99326061; PubMed=10398237;

RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,

RA Curtis M.;

RT "Carpenter-Waziri syndrome results from a mutation in XNP.";

RT Am. J. Med. Genet. 85:249-251(1999).

RN [16]

RP VARIANTS ATR-X GLU-175; 178-VAL--LYS-198 DEL; SER-190; PRO-219;

RP LEU-246 AND CYS-249.

RX MEDLINE=99219535; PubMed=10204841;

RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belougué J.,

```

Db      253  YEDERDQKEBNDYDTRSEASDGSSESVSFTD 285

RESULT 10
MP62_LYTP1
ID      MP62_LYTP1      STANDARD;      PRT;      411 AA.
AC      P91753;
DT      30-MAY-2000 (Rel. 39, Created)
DT      30-MAY-2000 (Rel. 39, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Mitotic apparatus protein P62.
OC      Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC      Echinoidae; Euechinoidae; Echinacea; Temnopleuroidea; Toxopneustidae;
OC      Lytechinus.
OX      NCBI_TaxId=7653;
RN      [1]
RS      SEQUENCE FROM N.A.
RX      MEDLINE=97166213; PubMed=9013612;
RY      Ye X., Sloboda R.D.;
RT      "Molecular characterization of p62, a mitotic apparatus protein
RT      required for mitotic progression.";
RL      J. Biol. Chem. 272:3606-3614(1997).
CC      -1- FUNCTION: REQUIRED FOR MITOTIC PROGRESSION. BINDS TO CHROMATIN.
CC      -1- SUBCELLULAR LOCATION: Nuclear.
CC      -1- PTM: PHOSPHORYLATED BY CAM-KINASE II IN VITRO.
CC      -1- SIMILARITY: BELONGS TO THE NUCLEOPLASMIN FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; U76750; AAB47481.1; -.
DR      InterPro; IPR004301; Nucleoplasmn.
DR      Pfam; PF03066; Nucleoplasmn; 1.
DR      SMART; S00001; Nucleoplasmn; 1.
DR      ProDom; PD00001; Nucleoplasmn; 1.
DR      ProTm; PT00001; Nucleoplasmn; 1.
FT      DOMAIN      147      156      POLY-GLU.
FT      DOMAIN      192      200      POLY-GLU.
FT      DOMAIN      262      267      POLY-GLU.
FT      DOMAIN      272      284      POLY-GLU.
FT      DOMAIN      293      300      POLY-ASP.
FT      DOMAIN      313      319      POLY-GLU.
SQ      SEQUENCE      411 AA; 46385 MW; D2DE04293C362254 CRC64;

Query Match      16.8%; Score 144.5; DB 1; Length 411;
Best Local Similarity 25.5%; Pred. No. 0.031;
Matches 42; Conservative 33; Mismatches 45; Indels 45; Gaps 8;

QY      3      QELTVSGSKRR-----RTGSLGNPSSGVNDEQNMNRVVEBQOOLROGBEHTARN 56
DB      155      EETFKGSPKRIYVKIAIYAKGRMGK--KGDELDD-----EDDEEEEEEESIIQTAKG 206
QY      57      GEVGVGVEPRPGQNDSDQ-----GTEENNRRFISVDEDSGNGQ----- 95
DB      207      KK-----RPAPSAKGPAAKLAVVDGTSKRKVPVGSVE--NGHAIIDDEDEBDEBDYKVG 260
QY      96      EEOGEDEBHAGEQDEBDEEEEMQDES---DFQSDSDGREDE 136
DB      261      DEEEEEATSGEEEEEDEEEEEEEDDEEMALGDDDDDEDEDEDE 305

RESULT 11
UBF1_RAT
ID      UBF1_RAT      STANDARD;      PRT;      764 AA.
AC      P25977; P25978;
DT      01-MAY-1992 (Rel. 22, Created)
DT      01-MAY-1992 (Rel. 22, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)

```

```

DE Nucleolar transcription factor 1 (Upstream binding factor 1) (UBF-1).
OS UBFR OR TOFUBR OR UBF-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (UBF1 AND UBF2) .
RX MEDLINE=91195316; PubMed=2014238;
RA O'Mahony D.J., Rothblum L.I.;
RT "Identification of two forms of the RNA polymerase I transcription
RT factor UBF."
RL Proc. Natl. Acad. Sci. U.S.A. 88:3180-3184(1991).
RN [2]
RP SUBUNIT.
RX MEDLINE=99199559; PubMed=10099786;
RA Riddsdale R.A., Semocok J.L., Larson D.E., Rothblum L.I., Hatauz G.;
RT "Topology of recombinant rat upstream binding factor."
RL Biochem. Cell Biol. 76:649-655(1998).
CC -1- FUNCTION: UBF RECOGNIZES THE RIBOSOMAL RNA GENE PROMOTER AND
CC ACTIVATES TRANSCRIPTION MEDIATED BY RNA POLYMERASE I THROUGH
CC COOPERATIVE INTERACTIONS WITH THE SPECIES-SPECIFIC FACTOR SL1. IT
CC BINDS SPECIFICALLY TO THE UPSTREAM CONTROL ELEMENT.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=UBF1; Synonyms=Long;
CC IsoId=P25977-1; Sequence=Displayed;
CC Name=UBF2; Synonyms=Short;
CC IsoId=P25977-2; Sequence=VSP_002195;
CC -1- SIMILARITY: Contains 6 HMG box domains.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1db-sib.ch/announce
CC or send an email to license@1db-sib.ch).
-----
CC EMBL; M61726; -; NOT_ANNOTATED_CDS.
DR EMBL; M61725; -; NOT_ANNOTATED_CDS.
DR PIR; A40439; A40439.
DR PIR; B40439; B40439.
DR HSSP; P07155; 1HMF.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box_5.
DR SMART; SMO0398; HMG_6
DR PROSITE; PS50118; HMG_BOX_2; 6.
RW Transcription regulation; DNA-binding; Activator; Nuclear protein;
KW Repeat; Alternative splicing.
FT DNA_BIND 112 180 HMG_BOX 1.
FT DNA_BIND 196 264 HMG_BOX 2.
FT DNA_BIND 298 362 HMG_BOX 3.
FT DNA_BIND 407 475 HMG_BOX 4.
FT DNA_BIND 482 549 HMG_BOX 5.
FT DNA_BIND 568 634 HMG_BOX 6.
FT DOMAIN 675 764 ASP/GLU/SER-RICH (ACIDIC).
FT VASPLIC 221 257 Missing (in isoform UBF2).
FT /FTId=VSP_002195.
SQ SEQUENCE 764 AA; 89437 MW; B9D37371615DB0534 CRC64;

Query Match 16.5%; Score 141.5; DB 1; Length 764;
Best local similarity 23.9%; Pred. No. 0.084;
Matches 42; Conservative 35; Mismatches 56; Indels 43; Gaps 6;

OY 1 MNDELLSVGSKRRRTGSLRGNPSSQGVDEQNNRVVEEQD-----QLR 46
Db : : : : : : : : : : : : : : : : : : : : : : : : : : :
593 LKERHVEIGSRWGI-----SSQKEHYKKLAEEDQRYKVKHLDLVKTSLSQDR 642
OY 47 QQEEHTRNR-----GEVYGVPRPRGQNDSCQGLLENNNRFTISVDEDSNGQEQEDE 102
: : : : : : : : : : : : : : : : : : : : : : : : : : :

```


OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 CX NCBI_Taxid:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99156230; PubMed=10048485;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 Miyajima N., Tanaka A., Kocani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 5:355-364(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21638749; PubMed=11780052;
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baggeley C.L.,
 Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
 Beasley O.P., Bird C.P., Blake S.E., Bridgeman A.M., Brown A.J.,
 Buck D., Burtill W.D., Butler A.P., Garder C., Carter N.P.,
 Chappman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 Grafham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 Kay M.P., Kimberley A.M., King A., Knights K., Laird G.K., Lawlor S.,
 Levaevscho M.H., Liversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 Marsh S.L., Martin S.L., McConachie L.J., McElay K., McMurtry A.A.,
 Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkhen R., Sims S.,
 Skuse C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 Whitehead S.L., Whitaker P., Willey D.L., Williams S.A.,
 Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RN [3]
 RP SEQUENCE OF 293-1121 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=9937452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,
 Tanaka A., Kocani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:197-205(1999).
 RN [4]
 RP SEQUENCE OF 424-1121 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=93078764; PubMed=1280325;
 RA Kim J.G., Hudson L.D.;
 RT "Novel member of the zinc finger superfamily: a C2-HC finger that
 RT recognizes a glia-specific gene.";
 RL Mol. Cell. Biol. 12:5632-5639(1992).
 RN [5]
 RP FUNCTION: BINDS TO THE PROMOTER REGIONS OF PROTEOLIPID PROTEINS
 CC -1- OF THE CENTRAL NERVOUS SYSTEM.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: MOSTLY IN DEVELOPING NERVOUS SYSTEM.
 CC -1- DOMAIN: CONTAINS 7 ZINC FINGERS OF THE C2HC CLASS ARRANGED IN
 CC TWO WIDELY SEPARATED CLUSTERS. THESE TWO DOMAINS OF DNA BINDING
 CC CAN FUNCTION INDEPENDENTLY AND RECOGNIZE THE SAME DNA SEQUENCE.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AB020642; BAA74858.1;
 CC EMBL: AL121581; CAC17005.1;
 CC EMBL: AB028973; BAA83002.1;
 CC EMBL: M96980; AAB59897.1;
 CC TRANSFAC: T04937;
 CC Genew: HGNC:7622; MYT1.
 CC GK: 001538;
 CC MIM: 600379;
 CC GO: GO:0005634; C:nucleus; NAS.
 CC GO: GO:0003700; F:transcription factor activity; NAS.
 CC GO: GO:0008270; F:zinc ion binding activity; NAS.
 CC GO: GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
 CC Interpro: IPR002515; ZnfC2HC.
 CC Pfam: PF01530; Zf-C2HC.7.
 CC Transcription regulation; Zinc-finger; DNA-binding; Nuclear protein;
 CC Repeat.
 CC ZN_FING 440 466 C2HC-TYPE.
 CC ZN_FING 484 510 C2HC-TYPE.
 CC ZN_FING 825 851 C2HC-TYPE.
 CC ZN_FING 874 895 C2HC-TYPE.
 CC ZN_FING 918 944 C2HC-TYPE.
 CC ZN_FING 971 997 C2HC-TYPE.
 CC CONFLICT 567 574 ATRPANA -> RHTGOLG (IN REF. 4).
 CC CONFLICT 616 616 C -> SKRPPKASPRSPSSSVTRSSSSAG (IN
 CC REF. 4).
 CC CONFLICT 837 837 A -> T (IN REF. 4).
 CC CONFLICT 943 943 G -> D (IN REF. 4).
 CC CONFLICT 957 958 AN -> TI (IN REF. 4).
 CC CONFLICT 1075 1075 R -> H (IN REF. 4).
 CC CONFLICT 1088 1089 DA -> VP (IN REF. 4).
 CC CONFLICT 1102 1102 D -> A (IN REF. 4).
 CC SEQUENCE 1121 AA; 122329 MW; DAFLR8CTD4BC01E CRC64;
 CC Query Match 17.0%; Score 146.5; DB 1; Length 1121;
 CC Best Local Similarity 24.7%; Pred. No. 0.062;
 CC Matches 40; Conservative 28; Mismatches 47; Indels 47; Gaps 5;
 CC
 CC QY 29 DEQGNRVYEEBOQ---QQLRQOEHTARNGEVGVPEPPGQNDQGGQLENNRRF 84
 CC DB 220 DAEVEVETTERSQDPCPSLEDAASESKQKGLSHE-----EEDDEEEEEE 273
 CC QY 85 ISVDESSNGQOEDEDEEAGQDEDEEEDDESD- -PDQ----- 127
 CC DB 274 ---DEEEEEEDEEEDEEEDEEEDDEEADPVITQEDTSHTSAQAPELRG 330
 CC QY 128 -----SDPSREDEHTHTSVTNGSSIVDL 152
 CC DB 331 PESPSKPEYSVAVEVRSDDDKEDTHSRKSTVTDESEMDM 372
 CC
 CC RESULT 8
 CC YGPO YEAST STANDARD; PRT; 1489 AA.
 CC ID YGPO YEAST
 CC AC P53115;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Hypochemical 171.5 kDa helix in NUT1-A02 intergenic region.
 CC GN YG150C OR G1880.
 CC OS Saccharomyces cerevisiae (Baker's yeast).
 CC OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycace.
 CC OX NCBI_Taxid=4932;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RX MEDLINE=96158061; PubMed=8585324;
 CC RA James C.M., Indge K.V., Oliver S.G.;
 CC RT "DNA sequence analysis of a 35 kb segment from Saccharomyces
 CC cerevisiae chromosome VII reveals 19 open reading frames including

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AL355013; CAB90146.1; -
 DR EMBL; Z98762; CAB11485.1; -
 DR EMBL; Z98599; CAB11250.2; -
 DR EMBL; D89140; BAA13802.1; -
 DR PIR; T38786; T38786;
 DR PIR; T42417; T42417;
 DR GeneDB; Spombe; SPAC4A8.16c; -
 DR InterPro; IPR000717; PCI.
 DR SMART; SM00088; PCI; 1.
 KW Initiation factor; Protein biosynthesis.
 FT DOMAIN 1 188 ASP/GLU/SER-RICH.
 FT CONFLICT 683 683 M -> I (IN REF. 2).
 SQ SEQUENCE 918 AA; 104367 MW; 43B70599409688BF CRC64;

Query Match 17.4%; Score 149.5; DB 1; Length 918;
 Best Local Similarity 24.0%; Pred. No. 0.034;
 Matches 36; Conservative 36; Mismatches 65; Indels 13; Gaps 4;

QY 16 GGSILRGNSQVDEQNNRVV-----EEQQOQLRQEEHTARNGEVGVPEPPGQGN 70
 DB 7 GGS--SDSDASVVSSEENRLTSSRLKQKQDSSSESESESSASSSESESE--ES 60
 QY 71 DSGOGLFENNRRISVDESSGQDEDEDEHAGDEDEDEDEDEDEDEDEDEDEDDSD 130
 DB 61 ESESEVEVPPKKKVAASEDESESESESESESESESESESESESESESESESESE 118
 QY 131 SSRDEHTNTSVNSSIVDLPHYQLSSP 160
 DB 119 ESE 148

RESULT 6
 UBF1 HUMAN STANDARD; PRT; 764 AA.
 AC P17480.
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Nucleolar transcription factor 1 (upstream binding factor 1) (UBF-1)
 DE (Anticentriolar NOR-90).
 GN UBF1 OR UBF1 OR UBF.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (UBF1).
 RX MEDLINE=90231434; PubMed=2330041;
 RA Jantzen H.M., Admon A., Bell S.P., Tjian R.;
 RT "Nucleolar transcription factor hUBF contains a DNA-binding motif
 RT with homology to HMG proteins";
 RT Nature 344:830-836(1990).
 RL [2]
 RP SEQUENCE FROM N.A. (UBF2).
 RX MEDLINE=92044316; PubMed=1940801;
 RA Chan E.K.L., Imai H., Hamel J.C., Tan E.M.;
 RT "Human autoantibody to RNA polymerase I transcription factor hUBF.
 RT Molecular identity of nucleolus organizer region autoantigen NOR-90
 RT and ribosomal RNA transcription upstream binding factor";
 RL J. Exp. Med. 174:1233-1244(1991).
 CC -1- FUNCTION: RECOGNIZES THE RIBOSOMAL RNA GENE PROMOTER AND ACTIVATES
 CC TRANSCRIPTION MEDIATED BY RNA POLYMERASE I THROUGH COOPERATIVE
 CC INTERACTIONS WITH THE SPECIES-SPECIFIC FACTOR SL1. IT BINDS
 CC SPECIFICALLY TO THE UPSTREAM CONTROL ELEMENT.
 CC -1- SUBUNIT: Homodimer.
 CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=UBF1; Synonyms=Long;
 CC IsoId=PI1480-1; Sequence=Displayed;
 CC Name=UBF2; Synonyms=Short;
 CC IsoId=PI1480-2; Sequence=VSP 002193;
 CC -1- SIMILARITY: Contains 6 HMG box domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; X53390; CAA37469.1; -
 DR EMBL; X53461; CAA37548.1; -
 DR EMBL; X56687; CAA40016.1; -
 DR PIR; S09318; S09318.
 DR PIR; S18193; S18193.
 DR PDB; 1K99; 04-DEC-02.
 DR TRANSFAC; T02900; -
 DR GeneW; HGNC:12511; UBF1.
 DR MIM; 600673; -
 DR GO; GO:0005730; C:nucleolus; TAS.
 DR GO; GO:0003701; F:RNA polymerase I transcription factor activity; TAS.
 DR GO; GO:0006356; P:regulation of transcription from Pol I prom. .; TAS.
 DR InterPro; IPR00910; HMG_12_box.
 DR Pfam; PF00505; HMG_box; 5.
 DR SMART; SM00398; HMG; 6.
 DR PROSITE; PSS0118; HMG_BOX_2; 6.
 KW Transcription regulation; DNA-binding; Activator; Nuclear protein;
 KW Repeat; Alternative splicing; 3D-structure.

FT DNA_BIND 112 180 HMG_BOX 1.
 FT DNA_BIND 196 264 HMG_BOX 2.
 FT DNA_BIND 298 362 HMG_BOX 3.
 FT DNA_BIND 407 475 HMG_BOX 4.
 FT DNA_BIND 482 549 HMG_BOX 5.
 FT DNA_BIND 568 634 HMG_BOX 6.
 FT DOMAIN 675 764 ASP/GLU/SER-RICH (ACIDIC).
 FT VARSPLIC 221 257 Missing (in isoform UBF2).
 FT /Ftrd-VSP 002193.
 SQ SEQUENCE 764 AA; 89406 MW; D4F08BE180E75D CRC64;

Query Match 17.4%; Score 147; DB 1; Length 764;
 Best Local Similarity 23.2%; Pred. No. 0.04;
 Matches 41; Conservative 41; Mismatches 57; Indels 44; Gaps 5;

QY 1 MNDELISVSGSKRRRTGSLGNPSSQVDEQNNRVVEEQOQ-----QLR 46
 DB 593 LKRMVWIGSRWRI-----SSQKEHYKRLAEQOKYKHLDMVSLSPQDR 642

QY 47 QQEEHTARNGEVV-----GVPEPPGQNDGQGLFENNRRISVDESSGQDEDEDE 102
 DB 643 AAYKEYISNKRKSMTKLRGNPKSSRTTLOSKSSEDDDE-----DDDDDEDEDEDE 698

QY 103 -----EYAGEDEDEDEDEDEDEDDFDSDSDSSRDEHTNTSVN 147
 DB 699 NGDSESDGDSSESSSESESEDEDEDEDEDEDEDEDEDEDEDESESSSSSS 755

RESULT 7
 MT1 HUMAN STANDARD; PRT; 1121 AA.
 ID MT1 HUMAN
 AC 001538; O94922; O9UPV2;
 DT 01-JUN-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Myelin transcription factor 1 (MT1) (MT1) (Proteolipid protein
 DE binding protein) (PLPB1).
 GN MT1 OR PLPB1 OR MT1 OR KIAA0835 OR KIAA1050.
 OS Homo sapiens (Human).

QY 3 QELLVSGSKRRRTG-----GSLRGNPSSQVDEQOMRVVEEEOOQLRQEEETAR 55
 DB 65 EEEVEEEERREERREEREGEGEGREAEEREEAEEREEAEEREEAEEREEAE 124
 QY 56 NGEVVGEPGPGGNDQOQOLENNNRFTSVDESSGNQOEBOEDEDEHAGEODEEDEE 115
 DB 125 EAE---AEEERAEEREEAEEREEAEEREEAEEREEAEEREEAEEREEAE 181
 QY 116 EEMDQESDDFDQSDSSREDE 136
 DB 182 EEEAEERAEERAEERAEERAE 202

RESULT 4
 CENB_CRIGR STANDARD; PRT; 606 AA.
 AC P48988;
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Major centromere autoantigen B (Centromere protein B) (CENP-B).
 GN
 OS Cricetus griseus (Chinese hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetidae;
 OC Cricetulus.
 NC NCBI_TaxID=10029;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=96254058; PubMed=8652663;
 RA Bejarano L.A., Valdivia M.M.;
 RT "Molecular cloning of an intronless gene for the hamster centromere antigen CENP-B";
 RL Biochim. Biophys. Acta 1307:21-25 (1996).
 CC -!- FUNCTION: INTERACTS WITH CENTROMERIC HETEROCHROMATIN IN CHROMOSOMES AND BINDS TO A SPECIFIC SUBSET OF ALPHOID SATELLITE DNA, CALLED THE CENP-B BOX. MAY ORGANIZE ARRAYS OF CENTROMERE SATELLITE DNA INTO A HIGHER ORDER STRUCTURE WHICH THEN DIRECTS CENTROMERE FORMATION AND KINETOCHORE ASSEMBLY IN MAMMALIAN CHROMOSOMES (BY SIMILARITY).
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- SIMILARITY: Contains 1 CENPB domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL: U20951; AAB06494.1; -.
 DR PIR: S70358; S70358.
 DR HSSP: P07199; 1BM6.
 DR InterPro: IPR004875; CENP-B.
 DR InterPro: IPR006600; CENPB.
 DR Pfam: PF04218; CENP-B_N; 1.
 DR Pfam: PF03184; DDE; 1.
 DR SMART: SM00674; CENPB; 1.
 KW Chromosomal protein; Nuclear protein; DNA-binding; Centromere.
 FT DNA BIND 1 125 BY SIMILARITY.
 FT DOMAIN 404 471 GLU-RICH (ACIDIC).
 FT DOMAIN 510 545 ASP/GLU-RICH (ACIDIC).
 SQ SEQUENCE 606 AA; 66407 MW; 63DDEE551D48E CRC64;

Query Match 17.4%; Score 149.5; DB 1; Length 606;
 Best Local Similarity 25.2%; Pred. No. 0.023;
 Matches 41; Conservative 22; Mismatches 47; Indels 53; Gaps 4;

QY 16 GGSLRGNPSS-----QVDEQOMRVVEEEOOQLRQEEETARNGEVGEPGPGND 71
 DB 16 GGSLRGNPSS-----QVDEQOMRVVEEEOOQLRQEEETARNGEVGEPGPGND 71

DB 390 GGGANNTTTSFKSGEEREEEREEEREEEREEEREEEREEEREEEREEEREE 445
 QY 72 SQOQOLENNNRFTSVDESSGNQOEBOEDE----- 103
 DB 446 GEEVEVEEBS-----DESEEEEREEEREEEREEEREEEREEEREEEREE 439
 QY 104 -----HAGEODEEDEEEMDQESDDFDQSDSSRED 135
 DB 500 EEAQCEPTLHLEGEDSDSDSEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 542

RESULT 5
 IF38_SCHPO STANDARD; PRT; 918 AA.
 AC O14164; O13885; P78791; Q9P6P4;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Probable eukaryotic translation initiation factor 3 93 kDa subunit (eIF3 p93).
 GN
 OS Schizosaccharomyces pombe (Fission Yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 NC NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A., Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S., Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M., Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A., Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G., Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagers K., James K., Jones L., Jones M., Leather S., McDonald S., McLean J., Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C., Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E., Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S., Shelton J., Simmonds M., Squares R., Squares S., Stevens K., Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S., Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B., Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S., Gabell C., Fuchs M., Fritz C., Holzer E., Moestl D., Hiltbert H., Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M., Eger P., Zimmermann W., Wedler H., Wambut R., Furnelle B., Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S., Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M., Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G., Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J., Dominguez A., Revelante J.L., Moreno S., Armstrong J., Forburg S.L., Cerritelli L., Lowe T., McCombie M.R., Paulsen I., Potashkin J., Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe";
 RL Nature 415:871-880 (2002).
 RN [2]
 RP SEQUENCE OF 571-918 FROM N.A.
 RC STRAIN=PR745;
 RA MEDLINE=98162722; PubMed=9501991;
 RA Yoshioke S., Kato K., Nakai K., Okayama H., Nojima H.;
 RT "Identification of open reading frames in Schizosaccharomyces pombe cDNAs";
 RL DNA Res. 4:363-368 (1997).
 CC -!- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF METHIONYL-TRNAI AND MRNA (BY SIMILARITY).
 CC -!- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS (BY SIMILARITY).
 CC -!- DOMAIN: CONTAINS 1 PCI DOMAIN.
 CC -!- SIMILARITY: BELONGS TO THE EIF38 FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

THIS PAGE BLANK (uspto)

Nature 344, 830-836, 1990

A:Title: Nucleolar transcription factor hNBF contains a DNA-binding motif with homology

A:Reference number: S09318; MUID:90231434; PMID:2330041

A:Accession: S09318

A:Molecule type: mRNA

A:Residues: 1-764 <JAN>

A:Cross-references: EMBL:X53461; NID:g37573; PIDN:CA37548.1; PID:g37574

A:Note: part of this sequence was confirmed by amino acid sequencing

C:Superfamily: Unassigned HMG box proteins; HMG box homology

C:Keywords: DNA binding; transcription regulation

F:103-184/Domain: HMG box homology <HMG3>

F:193-268/Domain: HMG box homology <HMG2>

F:294-366/Domain: HMG box homology <HMG3>

F:404-479/Domain: HMG box homology <HMG3>

F:482-553/Domain: HMG box homology <HMG3>

F:564-638/Domain: HMG box homology <HMG6>

Query Match 17.1%; Score 147; DB 2; Length 764;

Best Local Similarity 23.2%; Pred. No. 0.023; 57; Indels 44; Gaps 5;

Matches 41; Conservative 35; Mismatches 57; Indels 44; Gaps 5;

QY 1 MNGELLSVSGSKRRRTGSLKGNPSSQVDEQMRVVEEQO-----QLR 46

DB 593 LKRWVETIGSRWQRI-----SQSKHYKKLAEQOKYKHLDMVKSLSPODR 642

QY 47 QOEERTANRGEV---GVPRPGQNDSSQQLLENNRFTSVDESSGNQOEDEDE 102

DB 643 AAYKEYISNKKMTKLRGNPKSRTTLQSKSESEDEDE---DEDEDEDEDEDEDE 698

QY 103 -----EHAGEDEDEDEDEEMDQESDPFQSDSSREDEHTHTNSVTSS 147

DB 699 NGDSSEDDGSDSSSESEDESDGDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 755

RESULT 14

T42712

myelin transcription factor 1 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000

C:Accession: T42712

R:Kim, J.G.; Armstrong, R.C.; Robinsky, A.; Agoston, D.V.; Wiese, C.; Nagle, J.; Hudson, J. Neurosci. Res. 50, 272-290, 1997

A:Title: Myelin transcription factor 1 (Mtf1) of the oligodendrocyte lineage, along with

A:Reference number: Z20766; MUID:98038874; PMID:9373037

A:Accession: T42712

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1078 <KIM>

A:Cross-references: EMBL:AF04294; NID:g2209114; PID:g2209115; PIDN:AAC53456.1

C:Genetics:

A:Gene: Mtf1

A:Map position: 2

C:Keywords: DNA binding; transcription factor; zinc finger

Query Match 17.1%; Score 147; DB 2; Length 1078;

Best Local Similarity 26.8%; Pred. No. 0.031;

Matches 44; Conservative 27; Mismatches 43; Indels 50; Gaps 6;

QY 29 DEEQMRVVEEQOQLRQO-----EEHTANRGEVGVPRPGQNDSSQQLLENNR 83

DB 172 DVEEVLEIVTSERSQEPQSLKDMVSESSKQK-VLGHE-----EEGEEDEDEDE 221

QY 84 FTSVDESSG--NOEQEEDDEHAGQDEDEDEDE-----MDQSDDF 125

DB 222 EEDDEEEEGEGEGEEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 281

QY 126 D-----QSDDSREDEHTHTNSVTSSIVDL 152

DB 282 RGPBLSPPKREYVIVVRSDDDKDSDRSQKSAVTDSESEWDM 325

RESULT 15

JC5113

ribosomal transcription factor UBF2 - Chinese hamster

C:Species: Cricetus griseus (Chinese hamster)

C:Date: 31-Jan-1997 #sequence_revision 31-Jan-1997 #text_change 12-Feb-1999

C:Accession: JC5113

R:Bollivar, J.; Goenechea, L.G.; Grenett, H.; Pendon, C.; Valdivia, M.M.

Gene 176, 257-258, 1996

A:Title: Cloning and sequencing of the genes encoding the hamster ribosomal transcription

A:Reference number: JC5112; MUID:97075939; PMID:8918262

A:Accession: JC5113

A:Molecule type: mRNA

A:Residues: 1-727 <BOI>

A:Cross-references: GB:I42571

C:Comment: This factor binds to the ribosomal RNA gene's promoter and forms a stable pre

C:Superfamily: Unassigned HMG box proteins; HMG box homology

F:109-184/Domain: HMG box homology <HMG1>

F:257-329/Domain: HMG box homology <HMG3>

F:445-516/Domain: HMG box homology <HMG5>

Query Match 16.9%; Score 145.5; DB 2; Length 727;

Best Local Similarity 22.5%; Pred. No. 0.027;

Matches 40; Conservative 37; Mismatches 56; Indels 45; Gaps 5;

QY 1 MNGELLSVSGSKRRRTGSLKGNPSSQVDEQMRVVEEQO-----QLR 46

DB 556 LKRWVETIGSRWQRI-----SQSKHYKKLAEQOKYKHLDMVKSLSPODR 605

QY 47 QOEERTANR---GVVGVPRPGQNDSSQQLLENNRFTSVDESSGNQOEDEDE 102

DB 606 AAYKEYISNKKMTKLRGNPKSRTTLQSKSESEED---EEDDEDEDEDEDE 660

QY 103 EH-----AGQDEDEDEDEEMDQESDPFQSDSSREDEHTHTNSVTSS 148

DB 661 ENGDSSEDDGSDSSSESEDESDGDEDEDEDEDEDEDEDEDEDEDEDEDEDEDEDE 718

Search completed: December 10, 2003, 19:59:50

Job time : 22 secs

A:Cross-references: EMBL:U05314; NID:g476219; PID:g476220
 R:Rieger, M.; Mueller-Auer, S.; Schaefer, M.
 Submitted to the Protein Sequence Database, May 1996
 A:Reference number: S65202
 A:Accession: S65209
 A:Molecule type: DNA
 A:Residues: 1-802 <RIB>
 A:Cross-references: EMBL:Z73546; NID:g1370396; PID:g1370397; GSPDB:GN00016; MIPS:YPL190C
 A:Experimental source: strain S288C (AB972)
 R:Sumitomo, K.; Matsumoto, K.; Kornberg, R.D.; Reed, S.I.; Wittenberg, C.
 Mol. Gen. Genet. 248, 712-718, 1995
 A:Title: Dosage suppressors of the dominant G1 cyclin mutant CLN3-2: identification of a
 A:Reference number: S60122; MUID:96069710; PMID:7476874
 A:Accession: S60122
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-340, '1', 342-802 <SUG>
 A:Cross-references: GB:DJ7935; NID:g1235749; PID:g1235750
 C:Genetics:
 A:Gene: SGD:NAB3; NAB3; HMD1; MIPS:YPL190C
 A:Cross-references: SGD:S0006111
 A:Map position: 16L
 C:Superfamily: Yeast NAB3 protein; ribonucleoprotein repeat homology
 C:Keywords: nucleus; RNA binding
 F:331-396/Domain: ribonucleoprotein repeat homology <RNM1>

Query Match 17.7%; Score 152.5; DB 1; Length 802;
 Best Local Similarity 28.8%; Pred. No. 0.011;
 Matches 42; Conservative 29; Mismatches 48; Indels 27; Gaps 6;

QY 6 LSVGSKRRRTGSLRGNS-----SQVDEQNRVVEEQOQL-----RQEE 51
 DB 18 LSVGNSNSNE-NELMNNSSADGIEFDAPEREERAEEREEEOHELDVNDDEERDEE 75
 QY 52 HTARGEVGVGPRGGGNDSSQQLFENNRRFISVDESSGNOEEDDEHAGEODE 111
 DB 76 KGEERGVYINTE-----EEEREEHQKGGND---DDDNBEEREEEDDDDDDDDD 126
 QY 112 DEEEEMDQE--SDPDFQSDSSRED 135
 DB 127 DEEEEREEEGNDSSVGSDAED 152

RESULT 7

EDBEQ3 Immediate-early protein - saimiriine herpesvirus 1 (strain 11)

C:Species: saimiriine herpesvirus 1
 A:Note: host Saimiri sciureus (common squirrel monkey)

C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Jul-1999
 C:Accession: G36813; S20244
 R:Albrecht, J.

submitted to the EMBL Data Library, January 1992
 A:Description: Primary structure of the herpesvirus saimiri genome.
 A:Reference number: A36806

A:Accession: G36813
 A:Molecule type: DNA
 A:Residues: 1-407 <ALB>

A:Cross-references: GB:X64346; NID:G60320; PIDN:CAA5696.1; PID:G60394
 R:Albrecht, J.C.; Nicholas, J.; Biller, D.; Cameron, K.R.; Blesinger, B.; Newman, C.; W

J. Virol. 66, 5047-5058, 1992
 A:Title: Primary structure of the herpesvirus saimiri genome.

A:Reference number: A37309; MUID:92333688; PMID:1321287
 A:Contents: annotation; possible protein-coding frames
 A:Note: neither amino acid nor nucleotide sequence is given

R:Nicholas, J.; Cameron, K.R.; Honess, R.W.
 Nature 355, 362-365, 1992
 A:Title: Herpesvirus saimiri encodes homologues of G protein-coupled receptors and cycl

A:Reference number: S20243; MUID:92115001; PMID:1309943
 A:Accession: S20244

A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-407 <NTC>
 A:Cross-references: GB:S76368; NID:G243351; PIDN:AAH21116.1; PID:G243353

C:Genetics:
 A:Gene: 73
 C:Superfamily: herpesvirus immediate-early protein IE68
 C:Keywords: early protein

Query Match 17.4%; Score 150; DB 1; Length 407;
 Best Local Similarity 24.8%; Pred. No. 0.008;
 Matches 35; Conservative 32; Mismatches 64; Indels 10; Gaps 2;

QY 3 QELTVGSKRRRTG-----GSLRGNSSQVDEQNRVVEEQOQLRQEEHTAR 55
 DB 65 EEVEEERERRRRQEEEREGGEGGEGREAEAEAEAEAEAEAEAEAEAEAE 124
 QY 56 NGEVGVGPRGGGNDSSQQLFENNRRFISVDESSGNOEEDDEHAGEODEDEE 115
 DB 125 EAE---AEERAEERAEERAEERAEERAEERAEERAEERAEERAEERAE 181
 QY 116 EEMQESDDPDFQSDSSRED 136
 DB 182 EEEAEAEAEAEAEAEAEAEAE 202

RESULT 8

S70358 centromere protein B - Chinese hamster

C:Species: Cricetulus griseus (Chinese hamster)
 C:Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 05-Nov-1999

C:Accession: S70358
 R:Bejarano, L.A.; Valdivia, M.M.

Biotech. Biophys. Acta 1307, 21-25, 1996
 A:Title: Molecular cloning of an intronless gene for the hamster centromere antigen CENP

A:Reference number: S70358; MUID:96254058; PMID:8652663
 A:Accession: S70358

A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-606 <BE1>

A:Cross-references: EMBL:U20951; NID:G836955; PIDN:AAH0494.1; PID:G836956
 C:Keywords: DNA binding

Query Match 17.4%; Score 149.5; DB 2; Length 606;
 Best Local Similarity 25.2%; Pred. No. 0.013;
 Matches 41; Conservative 22; Mismatches 47; Indels 53; Gaps 4;

QY 16 GGSILRGNSSS-----QVDEQNRVVEEQOQLRQEEHTARNGEVGPRGGND 71
 DB 390 GCGANATITTSFSGEGEEEREEEREEEREEEREEEREEEREEEREEER 445
 QY 72 SQGQLFENNRRFISVDESSGNOEODEE----- 103
 DB 446 GEEVEVEEG-----DESDDEEEEREEEREEEREEEREEEREEEREE 499
 QY 104 -----HAGEODEDEEEEMDQESDDPDFQSDSSRED 135
 DB 500 EEAQPTLHFLGEGSDSDSDEEEDDEEEDDDDDDD 542

RESULT 9

T38786 translation initiation factor eif-3 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000

C:Accession: T38786; T38056
 R:Skellton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z21751
 A:Accession: T38786

A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 280-918 <SKR>

A:Cross-references: EMBL:Z28762; PIDN:CAH11485.1; GSPDB:GN00066; SPDB:SPAC4A8.16C
 A:Accession: T38056
 A:Status: preliminary; translated from GB/EMBL/DDAU
 A:Molecule type: DNA

R:Residues: 1-1115,1116-1485,1486-1657,1658-1722,1730-1948 <SCH>
A/Cross-references: EMBL:X07453

A: Molecule type: DNA
A: Residues: 1-802 <MIL>

A;Residues: 1-409 <STO>

THIS PAGE BLANK (USPTO)

PN WO200277036-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US08614.
 XX 21-MAR-2001; 2001US-277705P.
 XX (LEUNG/) LEUNG D W.
 XX Leung DW, Bergman PA, Lofquist A, Pietz GE, Tompkins CK;
 XX Waggoner DW;
 XX WPI; 2003-058367/05.
 XX
 XX Producing monodispersed preparation of polyanionic polymer for therapy,
 XX by expressing vector comprising ligation product of oligonucleotides
 XX encoding glutamate/aspartate residues in host cell and isolating the
 XX product -
 XX
 XX Example 6; Page 41; 74pp; English.
 XX
 XX The present invention describes a method (M) for producing a
 XX monodispersed preparation of a polyanionic polymer (PP) larger than
 XX 10 kD. (M) involves inserting into an expression vector (EV) a ligation
 XX product formed by ligating together oligonucleotides that encode
 XX glutamate/aspartate residues, expressing EV in a host cell, and
 XX isolating the protein product. (P) of EV, where (P) is PP and at least
 XX 80% of PP is approximately of the same molecular weight. Also described:
 XX (1) a recombinant fusion protein (I) comprising a polyanionic polypeptide
 XX and another polypeptide at either one end or at both ends of it; (2) a
 XX polyanionic polymer (II) conjugate comprising a polyanionic polymer and
 XX leucine, where the polyanionic polymer is polyglutamic acid or
 XX polyaspartic acid; (3) a vector (III) comprising a cassette which
 XX comprises a nucleotide sequence encoding a polyanionic polymer and at
 XX least one other nucleotide sequence, where the polyanionic polymer is
 XX polyglutamic acid or polyaspartic acid; (4) production of (I); (5) a cell
 XX (IV) comprising (III) or a vector that comprises a nucleotide sequence
 XX that encodes a polyanionic polymer that is larger than 10 kDa; and (6) a
 XX recombinantly-produced polyanionic polymer (V) that is of any molecular
 XX weight or is larger than 10 kD, and is conjugated to another protein.
 XX (I) is useful for treating a disease or ailment in an individual by
 XX administering (I) to the individual. (I) is also useful for delivering an
 XX effective amount of a pharmaceutically active agent, a therapeutic
 XX protein or a drug to a patient in need of it, or for diagnostic and
 XX testing or research purposes. AB222045 to AB222131 and ABP56374 to
 XX ABP6400 represent sequences used in the exemplification of the present
 XX invention.
 XX
 XX Sequence 187 AA;
 XX
 XX Query Match 19.4%; Score 167; DB 24; Length 187;
 XX Best Local Similarity 27.2%; Pred. No. 4.5e-07;
 XX Matches 31; Conservative 31; Mismatches 52; Indels 0; Gaps 0;
 XX
 XX 23 PSSQVDEQNMNRVVEEQOQLRQOEHTNARNGEVVGPVPRPGQNDSCQGLENNNN 82
 XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 XX Db 4 PEEEEESEEE 63
 XX
 XX 83 RFIIVDEDSGNQOEDEEHAGQDEDEEEEMQESDPDQSDSDSREDE 136
 XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 XX Db 64 EEE 117
 XX
 XX RESULT 10
 XX ABP56380
 XX ID ABP56380 standard; Peptide; 198 AA.
 XX
 XX AC ABP56380;
 XX
 XX 11-MAR-2003 (first entry)
 XX
 XX Polyanionic fusion protein product #1.
 XX
 XX DE

XX
 XX Polyanionic polymer; bioactivity; water solubility.
 XX Synthetic.
 XX WO200277036-A2.
 XX 03-OCT-2002.
 XX 21-MAR-2002; 2002WO-US08614.
 XX 21-MAR-2001; 2001US-277705P.
 XX (LEUNG/) LEUNG D W.
 XX Leung DW, Bergman PA, Lofquist A, Pietz GE, Tompkins CK;
 XX Waggoner DW;
 XX WPI; 2003-058367/05.
 XX
 XX Producing monodispersed preparation of polyanionic polymer for therapy,
 XX by expressing vector comprising ligation product of oligonucleotides
 XX encoding glutamate/aspartate residues in host cell and isolating the
 XX product -
 XX
 XX Disclosure; Page 28; 74pp; English.
 XX
 XX The present invention describes a method (M) for producing a
 XX monodispersed preparation of a polyanionic polymer (PP) larger than
 XX 10 kD. (M) involves inserting into an expression vector (EV) a ligation
 XX product formed by ligating together oligonucleotides that encode
 XX glutamate/aspartate residues, expressing EV in a host cell, and
 XX isolating the protein product. (P) of EV, where (P) is PP and at least
 XX 80% of PP is approximately of the same molecular weight. Also described:
 XX (1) a recombinant fusion protein (I) comprising a polyanionic polypeptide
 XX and another polypeptide at either one end or at both ends of it; (2) a
 XX polyanionic polymer (II) conjugate comprising a polyanionic polymer and
 XX leucine, where the polyanionic polymer is polyglutamic acid or
 XX polyaspartic acid; (3) a vector (III) comprising a cassette which
 XX comprises a nucleotide sequence encoding a polyanionic polymer and at
 XX least one other nucleotide sequence, where the polyanionic polymer is
 XX polyglutamic acid or polyaspartic acid; (4) production of (I); (5) a cell
 XX (IV) comprising (III) or a vector that comprises a nucleotide sequence
 XX that encodes a polyanionic polymer that is larger than 10 kDa; and (6) a
 XX recombinantly-produced polyanionic polymer (V) that is of any molecular
 XX weight or is larger than 10 kD, and is conjugated to another protein.
 XX (I) is useful for treating a disease or ailment in an individual by
 XX administering (I) to the individual. (I) is also useful for delivering an
 XX effective amount of a pharmaceutically active agent, a therapeutic
 XX protein or a drug to a patient in need of it, or for diagnostic and
 XX testing or research purposes. AB222045 to AB222131 and ABP56374 to
 XX ABP6400 represent sequences used in the exemplification of the present
 XX invention.
 XX
 XX Sequence 198 AA;
 XX
 XX Query Match 19.4%; Score 167; DB 24; Length 198;
 XX Best Local Similarity 27.2%; Pred. No. 4.8e-07;
 XX Matches 31; Conservative 31; Mismatches 52; Indels 0; Gaps 0;
 XX
 XX 23 PSSQVDEQNMNRVVEEQOQLRQOEHTNARNGEVVGPVPRPGQNDSCQGLENNNN 82
 XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 XX Db 11 PEEEEESEEE 70
 XX
 XX 83 RFIIVDEDSGNQOEDEEHAGQDEDEEEEMQESDPDQSDSDSREDE 136
 XX | : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
 XX Db 71 EEE 124
 XX
 XX RESULT 11
 XX ABG11270
 XX ID ABG11270 standard; Protein; 167 AA.
 XX
 XX

P1	Leung DW	Bergman PA,	Lofquist A,	Pietz GE,	Tompkins CK;
P1	Maggoner DW;				
XX					
XX					
DR	WPI; 2003-058367/05.				
XX					
PT	Producing monodispersed preparation of polyanionic polymer for therapy,				
PT	by expressing vector comprising ligation product of oligonucleotides				
PT	encoding glutamate/aspartate residues in host cell and isolating the				
PT	product -				
XX					
PS	Example 7; Page 46; 74pp; English.				
XX					
CC	The present invention describes a method (M) for producing a				
CC	monodispersed preparation of a polyanionic polymer (PP) larger than				
CC	10 kD. (M) involves inserting into an expression vector (EV) a ligation				
CC	product formed by ligating together oligonucleotides that encode				
CC	glutamate/aspartate residues, expressing EV in a host cell, and				
CC	isolating the protein product (P) of EV, where (P) is PP and at least				
CC	80% of PP is approximately of the same molecular weight. Also described:				
CC	(1) a recombinant fusion protein (I) comprising a polyanionic polypeptide				
CC	and another polypeptide at either one end or at both ends of it; (2) a				
CC	polyanionic polymer (II) conjugate comprising a polyanionic polymer and				
CC	leukin, where the polyanionic polymer is polyglutamic acid or				
CC	polyaspartic acid; (3) a vector (III) comprising a cassette which				
CC	comprises a nucleotide sequence encoding a polyanionic polymer and at				
CC	least one other nucleotide sequence, where the polyanionic polymer is				
CC	polyglutamic acid or polyaspartic acid; (4) production of (1); (5) a cell				
CC	(IV) comprising (III) or a vector that comprises a nucleotide sequence				
CC	that encodes a polyanionic polymer that is larger than 10 kDa; and (6) a				
CC	recombinantly-produced polyanionic polymer (V) that is of any molecular				
CC	weight or is larger than 10 kD, and is conjugated to another protein.				
CC	(1) is useful for treating a disease or ailment in an individual by				
CC	administering (I) to the individual. (I) is also useful for delivering an				
CC	effective amount of a pharmaceutically active agent, a therapeutic				
CC	protein or a drug to a patient in need of it, or for diagnostic and				
CC	testing or research purposes. AB22045 to AB22131 and ABP56374 to				
CC	ABP56400 represent sequences used in the exemplification of the present				
XX	invention.				
SQ	Sequence 197 AA;				
Query Match	19.5%; Score 168; DB 24; Length 197;				
Best Local Similarity	28.7%, Pred. NO. 3.9e-07;				
Matches 31; Conservative 30; Mismatches 47; Indels 0; Gaps 0;					
QY	29 DEQNRRVVEEEOQLRQEEHFRARNGEVGVPRPGGONDSSOGLNNNRNFI SVD 88				
Dd	76 EEE 135				
QY	89 EDSSGNQEQQEDDEEHAGQDEDEDEHEEMQDSDDPDQSDDS RDE 136				
Dd	136 EEEEEEEEEEEEEEGCGEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 163				
RESULT 8					
ABP56381					
ID	ABP56381 standard; Peptide; 186 AA.				
XX					
AC	ABP56381;				
XX					
DT	11-MAR-2003 (first entry)				
XX					
DE	Polyanionic fusion protein product #4.				
XX					
KW	Polyanionic polymer; bioactivity; water solubility.				
OS	Synthetic.				
XX					
PN	WO200277036-A2.				
XX					
FD	03-OCT-2002.				
XX					
PF	21-MAR-2002; 2002WO-US08614.				

XX	21-MAR-2001; 2001US-277705P.
PR	(LEUNG/) LEUNG D W.
XX	
PA	Leung DW, Bergman PA, Lofquist A, Pietz GE, Tompkins CK,
PI	Maggoner DW;
XX	
DR	WPI; 2003-058367/05.
XX	
PS	Example 4; Page 35; 74pp; English.
XX	
CC	The present invention describes a method (M) for producing a
CC	monodispersed preparation of a polyanionic polymer (PP) larger than
CC	10 kD. (M) involves inserting into an expression vector (EV) a ligation
CC	product formed by ligating together oligonucleotides that encode
CC	glutamate/aspartate residues, expressing EV in a host cell, and
CC	isolating the protein product (P) of EV, where (P) is PP and at least
CC	80% of PP is approximately of the same molecular weight. Also described:
CC	(1) a recombinant fusion protein (I) comprising a polyanionic polypeptide
CC	and another polypeptide (II) at either one end or at both ends of it; (2) a
CC	polyanionic polymer (II) conjugate comprising a polyanionic polymer and
CC	leukine, where the polyanionic polymer is polylutamic acid or
CC	polyspartic acid; (3) a vector (III) comprising a cassette which
CC	comprises a nucleotide sequence encoding a polyanionic polymer and at
CC	least one other nucleotide sequence, where the polyanionic polymer is
CC	polylutamic acid or polyspartic acid; (4) production of (I); (5) a cell
CC	(IV) comprising (III) or a vector that comprises a nucleotide sequence
CC	that encodes a polyanionic polymer that is larger than 10 kDa; and (6) a
CC	recombinantly-produced polyanionic polymer (V) that is of any molecular
CC	weight or is larger than 10 kD, and is conjugated to another protein.
CC	(I) is useful for treating a disease or ailment in an individual by
CC	administering (I) to the individual. (I) is also useful for delivering an
CC	effective amount of a pharmaceutically active agent, a therapeutic
CC	protein or a drug to a patient in need of it, or for diagnostic and
CC	testing or research purposes. AB22045 to AB22131 and ABP6374 to
CC	ABP6400 represent sequences used in the exemplification of the present
CC	invention.
CC	
XX	
Sequence	186 AA;
XX	
Query Match	19.4%; Score 167; DB 24; Length 186;
Best Local Similarity	27.2%; Prod. No. 4.4e-07;
Matches	31; Conservative 31; Mismatches 52; Indels 0; Gaps 0;
QY	23 PSSQVDEQOMRVVEEQOOLRQOEHTTANGVEVGVPRPGQONDQOQLENNN 82
DB	11 PEEEEEVEEE 70
QY	83 RFLSVDEDSGNGQEQEDDEHAGQDEEEFEEMDQESDDPDDQSDSRREF 136
DB	71 EEE 124
RESULT 9	
ABP6384	
ID	ABP6384 standard; Peptide; 187 AA.
XX	
AC	ABP6384;
XX	
DT	11-MAR-2003 (first entry)
XX	
DE	Polyanionic fusion protein product #4.
XX	
KW	Polyanionic polymer; bioactivity; water solubility.
XX	
OS	Synthetic.
XX	

PT	diagnostics, forensics, gene mapping, identification of mutations
FT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
PS	Claim 20; SEQ ID No 35713; 103bp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC	and genome mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 144 AA;
Query Match	19.9%; Score 171; DB 22; Length 144;
Best Local Similarity	30.8%; Pred. No. 1.5e-07;
Matches	33; Conservative 28; Mismatches 42; Indels 4; Gaps 1
Oy	30 EEQNRRVVEEEOQQOLRQOEDEHTANGEVGVGPVRPGQNSOQQLBENNRRFISYDE 89
Dd	2 EEKEKEKEEEKEEEEEKEEEEKEDEGE---EGGGGGEKGEBGEBEEEEEEEEEE 57
Oy	90 DSGNQEOQEDEDENHAGEODEDEDEEEMDOESDIPQSDDSREDE 136
Dd	58 EEE 104
RESULT 6	
ABG11263	
ID	ABG11263 standard; Protein; 172 AA.
XX	
AC	ABG11263;
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	Novel human diagnostic protein #11254.
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
PD	11-OCT-2001.
PF	30-MAR-2001; 2001MO-US08631.
PR	31-MAR-2000; 2000US-0540217.
PR	23-AUG-2000; 2000US-0649167.
PA	(HYSE-) HYSEQ INC.
PI	Dzmanac RT, Liu C, Tang YT;
Pt	WPI; 2001-639362/73.
DR	N-PSDB; AAS75450.

XX	New isolated polynucleotide and encoded polypeptides, useful in
PT	diagnostics, forensics, gene mapping, identification of mutations
PT	responsible for genetic disorders or other traits and to assess
PT	biodiversity -
XX	
PS	Claim 20; SEQ ID No 41622; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC	polynucleotide chain reaction (PCR) primers, oligomers, and for chromosome
CC	and genome mapping, and in recombinant production of (II). The
CC	polynucleotides are also used in diagnostics as expressed sequence tags
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
CC	to restore normal activity of (II) or to treat disease states involving
CC	(II). (II) is useful for generating antibodies against it, detecting or
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
CC	a food supplement. (II) and its binding partners are useful in medical
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
CC	disorders involving aberrant protein expression or biological activity.
CC	The polypeptide and polynucleotide sequences have applications in
CC	diagnostics, forensics, gene mapping, identification of mutations
CC	responsible for genetic disorders or other traits to assess biodiversity
CC	and to produce other types of data and products dependent on DNA and
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
CC	diagnostic amino acid sequences of the invention.
CC	Note: The sequence data for this patent did not appear in the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SO	Sequence 172 AA:
XX	
Query Match	19.6%; Score 168.5; DB 22; Length 172;
Best Local Similarity	24.3%; Pred. No. 3e-07;
Matches	35; Conservative 40; Mismatches 48; Indels 21; Gaps 2
Qy	1 MNQELLISGSKRRRTGSLRGNPSSGVDERQNMNVVEEQOQOLRQOEHTTANGEVY 60
Db	42 VGOVFLALIQPRRS-----KTPSKKKKEEEEEEEEEEEEEEEEEEE----- 88
Qy	61 GVEPRPGQNDSSQOGLENNRRFISVDESSGNQROEDENHAGDEDEDEEMDQ 120
Db	89 -----EE 140
Qy	121 ESDPDFQGDSSRDEDEHTINSVT 144
Db	141 EEEEEEEEEEEEEEEGSKINNLIS 164
XX	
RESULT 7	
ABP56385	
ID	ABP56385 standard; Peptide; 197 AA.
XX	
AC	ABP56385;
XX	
DT	11-MAR-2003 (first entry)
XX	
DE	Polyanionic fusion protein product #5.
XX	
KM	Polyanionic polymer; bioactivity; water solubility.
XX	
OS	Synthetic.
XX	
PN	WO200277036-A2.
XX	
PD	03-OCT-2002.
XX	
PF	21-MAR-2002; 2002WO-US08614.
XX	
PR	21-MAR-2001; 2001US-27705P.
XX	
PA	(LEUNG/) LEUNG D W.
XX	

(HYSE-) HYSBQ INC.
PI Dmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSTB; AAS75467.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 41639; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG00010-ABG037 represent novel human diagnostic amino acid sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SEQ Sequence 254 AA.

Query Match 20.3%; Score 174.5; DB 22; Length 254;
Best Local Similarity 25.0%; Pred. No. 1.4e-07;
Matches 39; Conservative 38; Mismatches 44; Indels 35; Gaps 4;

Yr 9 GSKRRRTGSLRGNDSSGVDEQNRRVVEEQOOLRQOEHEHTARRNGEVGVERPRFG 68
Db 55 GEERRRK--EKKGEEEEEE-----EEEEEEEEEEEEEEEEEEEE-----95
Yr 69 QNDSOGOLEENNRFIVDEDSNOBOEDDEHAEODEDEEEEMDQESDDPDQS 128
Db 96 --EEEEEEEE-----EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 146
Yr 129 DDSGREDEHTNTSVTSSSIYD-----LFPVOL 157
Db 147 EEEEEEGLSTIKWTSTKTLSDINGFKPLVPVHI 182

RESULT 4
AAR70491
ID AAR70491 standard; Protein; 562 AA.
AC AAR70491;
DT 19-DEC-1995 (first entry)
XX
XX leucocytozoan protozoa structural protein epitope.
DE leucocytozoan protozoa, structural protein; epitope; vaccine; fowl,
KM leucocytozanosis; treatment.
XX
OS leucocytozoan protozoa sp.
XX
PN JP07089995-A.
PD 04-APR-1995.

[illegible]

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS Claim 20; SEQ ID No 41624; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 194 AA;
 SQ

Query Match 20.3%; Score 175; DB 22; Length 194;
 Best Local Similarity 29.8%; Pred. No. 9.1e-08;
 Matches 34; Conservative 30; Mismatches 42; Indels 8; Gaps 2;

QY 23 PSSGVNDEQNNRVVEEEOOQLNQOEHEHARNGEVGVRPGGONDGOGGLENNN 82
 DB 28 PSEKEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEE 84
 QY 83 RFISVDEDSGNOEEDDEEHAGODEDEEEMDQESDPQSDPSRDE 136
 DB 85 -----EEEBEGGEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 133

RESULT 2
 ID ABG06966
 AC ABG06966;
 DT 13-FEB-2002 (first entry)
 DE Novel human diagnostic protein #6957.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 PA (HYSE-) HYSEQ INC.
 PI Drmanac RT, Liu C, Tang YT;
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS71153.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity
 PS Claim 20; SEQ ID No 37325; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 254 AA;
 SQ

Query Match 20.3%; Score 174.5; DB 22; Length 254;
 Best Local Similarity 25.0%; Pred. No. 1.4e-07;
 Matches 39; Conservative 38; Mismatches 44; Indels 35; Gaps 4;

QY 9 GSKRRRTGSLRGNPSSGVNDEQNNRVVEEEOOQLNQOEHEHARNGEVGVRPGG 68
 DB 55 GEEERRR--EKKGBEEEESEEESEEESEEESEEESEEESEEESEEESEEE 95
 QY 69 QNDSOQGLENNNRFISVDEDSGNOEEDDEEHAGODEDEEEMDQESDPDQS 128
 DB 96 --EEEEEESEEE-----EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE 146
 QY 129 DSSGREDEHTHNSVTNSSIVD-----LPHQL 157
 DB 147 EEEEEEEEGSLIKYMTSTKTLSDINGPEKLLPVHPI 182

RESULT 3
 ID ABG11280
 AC ABG11280;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #11271.
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 PN WO200175067-A2.
 PD 11-OCT-2001.
 PF 30-MAR-2001; 2001WO-US08631.
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2003, 19:52:45 ; Search time 40 Seconds
(without alignments)
662.683 Million cell updates/sec

Title: US-10-245-618-4
Sequence: 1 MNGELSVGSKRRRTGSLR.....SYDLPVHQLSPPYKTKK 167

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863
Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1980.DAT.*
 - 2: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1981.DAT.*
 - 3: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1982.DAT.*
 - 4: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1983.DAT.*
 - 5: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1984.DAT.*
 - 6: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1985.DAT.*
 - 7: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1986.DAT.*
 - 8: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1987.DAT.*
 - 9: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1988.DAT.*
 - 10: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1989.DAT.*
 - 11: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1990.DAT.*
 - 12: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1991.DAT.*
 - 13: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1992.DAT.*
 - 14: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1993.DAT.*
 - 15: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1994.DAT.*
 - 16: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1995.DAT.*
 - 17: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1996.DAT.*
 - 18: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1997.DAT.*
 - 19: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1998.DAT.*
 - 20: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA1999.DAT.*
 - 21: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2000.DAT.*
 - 22: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2001.DAT.*
 - 23: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2002.DAT.*
 - 24: /SIDSL/gcgdata/geneeq/geneeq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	175	20.3	194	22	ABG11265
2	174.5	20.3	234	22	ABG06566
3	174.5	20.3	254	22	ABG11280
4	172.5	20.1	562	16	AA070491
5	171	19.9	144	22	ABG05354
6	168.5	19.6	172	22	ABG11263
7	168	19.5	197	24	ABP56385
8	167	19.4	186	24	ABP56381
9	167	19.4	187	24	ABP56384

10	167	19.4	198	24	ABP56380	Polyanionic fusion
11	165.5	19.2	167	22	ABG11270	Novel human diago
12	165	19.2	360	17	AAW03627	Human follicle sti
13	164	19.1	412	21	AAW03626	Human thyrotropin
14	163.5	19.0	1162	21	AAW96255	Kaposi's sarcoma-a
15	163.5	19.0	1162	21	AAW58500	HN78 ORF 73 protei
16	163.5	19.0	1162	21	AAW62331	Amino acid sequenc
17	163.5	19.0	1162	23	ABG05621	Kaposi's sarcoma-a
18	161	18.7	141	22	ABG04739	Novel human diago
19	161	18.7	181	22	ABP56382	Polyanionic fusion
20	160.5	18.7	1132	17	AAW97866	Chicken leucocyte
21	160	18.6	508	22	ABG04359	Novel human diago
22	159	18.5	186	22	ABG06961	Novel human diago
23	158	18.4	169	22	ABG06954	Novel human diago
24	158	18.4	204	22	ABG05352	Novel human diago
25	157.5	18.3	199	22	ABG05362	Novel human diago
26	157	18.3	217	22	ABG05366	Novel human diago
27	154	17.9	147	22	ABG11269	Novel human diago
28	153.5	17.8	1174	22	AAW85039	Novel human diago
29	152.5	17.7	802	23	ABG93215	S. cerevisiae BAX-
30	152	17.7	207	22	ABG05360	Shrimp white spot
31	151	17.6	235	22	ABG11274	Novel human diago
32	150.5	17.5	205	22	ABG03977	Novel human diago
33	150	17.4	682	21	AAW46506	Novel human diago
34	150	17.4	724	21	AAW46505	Arabidopsis thalia
35	150	17.4	778	21	AAW46504	Arabidopsis thalia
36	149.5	17.4	425	22	ABG06903	Novel human diago
37	149	17.3	231	22	ABG26553	Novel human diago
38	149	17.3	237	22	ABG06965	Novel human diago
39	149	17.3	237	22	ABG11276	Novel human diago
40	148.5	17.3	108	22	ABG05368	Novel human diago
41	148.5	17.3	165	22	ABG05368	Novel human diago
42	147.5	17.2	207	21	AAW4672	Gene 33 human secr
43	147	17.1	727	23	ABW97597	Novel human protei
44	147	17.1	745	22	ABG22102	Novel human diago
45	146	17.0	191	22	ABG03929	Novel human diago

ALIGNMENTS

RESULT 1	ABG11265	ABG11265 standard; Protein; 194 AA.
ID	ABG11265;	
AC	ABG11265;	
XX		
DT	18-FEB-2002 (first entry)	
XX		
DE	Novel human diagnostic protein #11265.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001WO-US08631.	
XX		
PR	31-MAR-2000; 2000US-0540217.	
XX		
PR	23-AUG-2000; 2000US-0649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Dmanac RT, Liu C, Tang YT;	
XX		
DR	WPI; 2001-639362/73.	
XX		
DR	N-PSDB; AAS75452.	
XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in	

THIS PAGE BLANK (uspto)

Job time : 5028 secs

KKIKHOWYRADNEFKVALLKHLKODATRSIVFVKRERVHELAETLRLAGINNC
YLSEMAQIKRNEGIKRLTDGRVNVAVTAARAGDIPVSHVINDFMRSGDTYLA
RIETGAGKRGTAISVEAHDHLLIKIRYIEPKAVINDEIETPTTAPSEKLTG
KPSKVIYAKRAEKKKEKEKPKYKRYRPTKNIKKRRKSGTKMQDSSE"

6365. .7102
/gene="yfiC"
/locus_tag="t0268"
6365. .7102
/gene="yfiC"
/locus_tag="t0268"
/note="Corresponds to STY2835 from Accession AL513382:
Salmomella typhi CT18"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA067993.1"
/db_xref="GI:29136427"
/translation="MSQSGVLRNNGFTFKQFVANDRCAMKVTGDIILGAMAPVAD
VKRILDIGTSGGLALMLAOTDSDVDAVELDAGAMQAEVNAHSPWPIVHT
DDIORMAPROTVREDLISNPYEPGEVSTQREARVATLIDHOTLLAADCT
EDGFECVLEPGIAGFTQALNMGWHLRLRTDVAERAEALPHRVLLAFSPQAGECS
DRIVRSGDNYSESYTALTOAPFLFM"

complement (7087. .8709)
/gene="nadB"
/locus_tag="t0269"
complement (7087. .8709)
/gene="nadB"
/locus_tag="t0269"
/note="quinolinate synthetase B; corresponds to STY2834
from Accession AL513382: Salmomella typhi CT18"
/codon_start=1
/transl_table=11
/product="L-aspartate oxidase"
/protein_id="AA067994.1"
/db_xref="GI:29136428"
/translation="MMTTPELSCDVLLIGSAGALSIALRLAKKHVIVLSKPVSE
STFVAGGIAVEDETDSIAHVEDTLIAAGICDRHAEVFNATCVOMIDGV
LPTHVQNGKESYHLRREGSHSRITLHAADATGKEVETLVSRAQNHQVLEERS
NAVDLISDKIGLPGRVYVGAMSHNRKEMVETCAKSVIATGCA SKVYQYTTDD
ISSGDIAMAMRACRYANLEPNQFHTALYHQAARFLTEALRSGAYLKRPDSR
FMDVDERSGLAPDVARAIIDEMKRLGADCMFLDSKRPDVFVRQHPMTYAKLID
LMDLTKERIPIVPAAHYTCGVVVDYGRDVGALVAIEVSYTGIAHGRNVAASL
LECIYVGSAMADIDRMPSVHSDVLPANDESEVENADERVVIOHNMHELRLIMDY
VGIVRTKRLERALARITMLQOIDEVYAPRSNNLELRNLVQVAELIVRCAMRK
ESRGLHTLDYPOQLAESPSILSPFLPHNR"

9134. .9709
/gene="trpB"
/locus_tag="t0270"
9134. .9709
/gene="trpB"
/locus_tag="t0270"
/note="sigma-24; corresponds to STY2833 from Accession
AL513382: Salmomella typhi CT18"
/codon_start=1
/transl_table=11
/product="RNA polymerase sigma-E factor"
/protein_id="AA067995.1"

Alignment Scores:
Pred. No.: 7.58e-130 Length: 300431
Score: 1831.00 Matches: 365
Percent Similarity: 99.19% Conservative: 4
Best Local Similarity: 98.12% Mismatches: 0
Query Match: 98.12% Indels: 0
Gaps: 0

US-09-921-992-50 (1-372) x AE016835 (1-300431)

Qy 1 MethisanginalaProleGlnArgGlySerThrArgIleTyValGlyAsnVal 20
Db 83116 ATGCATACCAAGGCTCCGATTCAACGTAGAAATCGACACATATTTCGTTGGGAAGTGTG 83175

Qy 21 ProileGlyaspGlyAlaProileAlaValGlnserMetThrAenThrArgThrThrasp 40

|||||
Db 83176 CCATTGGCGATGGCCCTATCGCGCTGCATGACTACACGGCTACCTGAC 83235
Qy 41 ValGluAlaThyValAsnGlnIleTyValAlaLeuGluArgValGlyAlaSerIleValArg 60
Db 83236 GTGGAAGCAGCAGGTTATCATGATCAAGGCGCTGGAGCGCGGTTGGCCAGATATATGTTGGT 83295
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheTyLeuIleTySerGlnGlnValAsn 80
Db 83296 GTTTCCTGACCGACATGACGCGCGCGCGGAGGCTTACAGCTTATCAACACGAGGTTAGC 83355
Qy 81 ValProLeuValAlaAspIleHisPheAspTyArgIleAlaLeuTyValAlaGluTyArg 100
Db 83356 GTCCCGCTGGTGGCCGATATTCATTCCTCGACTACCGCATTCGCGTGAAGTAGCGGAATAC 83415
Qy 101 GlyValAspCysAlaLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgMet 120
Db 83416 GGGGTGACTGCTGTGGATTTAAACCCGCAATATGATGTAACGAAACGCAATTCGCATG 83475
Qy 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
Db 83476 GTGGTGAATCGCGCGCGCGCAAAAACATCCCTATCCGATCGCGGTAATGCGGATCG 83535
Qy 141 LeuGluTyAspLeuGlnGluTyArgIleGluProThrProGlnAlaLeuLeuGluSer 160
Db 83536 CTGGAATAAGATCTGCAGAAAATAACGAGACCGACCGCGCGCTGCTGGAATCA 83595
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheTyValSerVal 180
Db 83596 GCATCGCGCATGTCATCATCTCGATCTGCTCAACTTGTATCAAGTTGATCAAGGACCGTA 83655
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyArgLeuLeuAlaTySerGlnIleAsp 200
Db 83656 AAGGCTCGATATATTCCTTCCCTGAGTCTTATTCGTTGGCTAAACGATCGAT 83715
Qy 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValArgSerGlyAlaValLysSer 220
Db 83716 CACCGCTGATCTGGGATCAACGAAAGCGGGCGCGGTAGCGGGCGGTTAAGTGC 83775
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlyIleGlyAspThrLeuArgValSerLeu 240
Db 83776 GCATCGGTTAGTCTGCTGTTATGGAAGGATCGCGACCGCTGCGCGTATCGTTG 83835
Qy 241 AlaAlaAspProValGluGluIleTyValGlyPheAspIleLeuTySerLeuArgIle 260
Db 83836 CGCGCGACCGGTGGAAGGATCAAGTCGTTGATATCTGAAGTCGCTGCGTATT 83895
Qy 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnIlePheAspVal 280
Db 83896 CGCGCGCGCGGATCAATTTCATCGCTGCGCGACCTGTTCTGTCAGAGGTTGACGTT 83955
Qy 281 IleGlyThrValAsnAlaLeuGlnGluAlaGluLeuGluAspIleIleThrProMetAspVal 300
Db 83956 ATCGGACCGGTGAACCGCTGAGCGAGCGCTGGAAGATATCATCCCCGATGAGCGTC 84015
Qy 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
Db 84016 TCATCATCGTTGTGTGTGAACGAGCGCGGGAAGCGCTGATTCGACGCTGGGCGTG 84075
Qy 321 ThrGlyIleAsnLysLysSerGlyLeuTyArgLysPglValArgLysAspArgLeuAsp 340
Db 84076 ACGGGCGGCAATTAAGAAAGCGGCTGTATGAAGAGCGGCTGCTGAAGGACACATCGAT 84135
Qy 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
Db 84136 AACGACGATATGATTCGCGACGCTTGATCCGATTCGCGCGGAAGTAAATGTAACATTGAT 84195
Qy 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
Db 84196 GAAGCGCTCGATTCACGTCGTCAGGTTGAAAA 84231

RESULT 15
AE016835 300431 bp DNA linear BCT 21-MAR-2003
LOCUS
DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 2 of
16 of the complete genome.
ACCESSION
VERSION AE016835 AE014613
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. .300431
/organism="Salmonella enterica subsp. enterica serovar
Typhi Ty2"
/mol_type="genomic DNA"
/strain="Ty2"
/serovar="Typhi"
/sub_species="enterica"
/db_xref="taxon:209261"
complement (67. .486)
/gene="yfiG"
/locus_tag="t0261"
complement (67. .486)
/gene="yfiG"
/locus_tag="t0261"
complement (67. .486)
/gene="yfiG"
/locus_tag="t0261"
/note="corresponds to STY2842 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="thioredoxin 2"
/protein_id="AA067987.1"
/db_xref="GI:29136421"
/translation="MNTVCTGCOAIRIPEDRLQDAKCGRGHELFDGEVINATGET
LDRLKDDLPVVIDPNAWPCGRCNRPAPRDEBRSKGVKVTNTERELRSARF
GHSIPWIKIFKRGQVVDMLNGAVPAPFDSWLEAL"
690. .1727
/gene="yfiF"
/locus_tag="t0262"
690. .1727
/gene="yfiF"
/locus_tag="t0262"
/note="corresponds to STY2841 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative RNA methyltransferase"
/protein_id="AA067988.1"
/db_xref="GI:29136422"
/translation="MNDLKNKSGKYVMVRSDDSDKRTNPRTKGGGGRPAKST
DGRPARDERNNOSDRKHETSPMRTVSAPDEPBEKVDHGSGSKSFIDEVBER
ORAEFRVYGENAQQALFOSRPAIVRAWFGVTPRFKALEMWAANKAYIVVDEA
ELKASGTEHGGVCFULIKRNGITVKQWYKQADDCVLALEDVANPHNLGSMWSC
AHGCVGVVQVADALLESQAIRFAGGAHVQPIIGESIVDVLDDPFGAGYTVVTS
SDGQALFSTTLPEKMWLVYGREYDYLPEAAREPDDLVCVINGTVESLNVSVATGV
LLEWWRONKA"
complement (1843. .2532)

CDS
/gene="ung"
/locus_tag="t0263"
complement (1843. .2532)
/gene="ung"
/locus_tag="t0263"
/note="corresponds to STY2840 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="uracil-DNA glycosylase"
/protein_id="AA067989.1"
/db_xref="GI:29136423"
/translation="MATELTHWDVLADEKQCFYINTLHTVAGERQSGTTPYPPQKV
FNAFRTBLGVKVVIGODPYHQAAGLFSVAPGJAPPPLVMWKELEASIPG
FYRPAHGYESNAROGVLLNLVTVRAGASHASIGMETFTDKVTSLINOREGV
FLMGSHACKGAIIIDPQHNLIKAPHPSPLSAHKGFPCCNFPALTNQLEQHGERTI
DNTPVLPARSE"
2851. .3234
/gene="yfiD"
/locus_tag="t0264"
2851. .3234
/gene="yfiD"
/locus_tag="t0264"
/note="corresponds to STY2839 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA067990.1"
/db_xref="GI:29136424"
/translation="MTGQITKANDDLNSFWLDSKGBARCTIVAKGPAEDBEV
AVSKGEIYREIPEWKEVEVGGQHLNVVLRRETDVAVKPEKXPOLITRISG
VAVRPNSLTPBOQDVIAFTFESL"
complement (3296. .3883)
/gene="yfiK"
/locus_tag="t0265"
complement (3296. .3883)
/gene="yfiK"
/locus_tag="t0265"
/note="corresponds to STY2838 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative membrane protein"
/protein_id="AA067991.1"
/db_xref="GI:29136425"
/translation="MTPMLLSAFWTYTLTALTPSPNNILASAAITAHGFRQSIRVLA
GMSLGLVVMLCAGIASLAVIDPAIHLISWGAAYILMLAWKIATSPADAEKVP
KEVGFWSFGLQFVNVKIIIGITLSTFVLPOALNWNVIGVSILALIGTFGNVCW
ALAGHLFOARFHHYGRQLNIIALLLVCAVARIFV"
3986. .4868
/gene="yfiE"
/locus_tag="t0266"
/note="corresponds to STY2837 from Accession AL513382:
Salmonella typhi CT18"
/pseudo
complement (4901. .6235)
/gene="srmb"
/locus_tag="t0267"
complement (4901. .6235)
/gene="srmb"
/locus_tag="t0267"
/note="corresponds to STY2836 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="ATP-dependent RNA helicase Srmb"
/protein_id="AA067992.1"
/db_xref="GI:29136426"
/translation="MTVTFSLELDESLLDALQDKGFTRPYALQAAVLPALDGRDV
LGSAPGTGCTAAVYLPALQHLDPFRKSGPRLITLPTRELMAQVADHARELAKH
THLDIATITGVAVNMHAEFSENDIVVATGTGRLLOYIKENPFCRAVETLILDEAD
RMLDMPADIDIHINGETRMKQTMLEFATLEGDALXDPARELLDDPVEVSANPSTRE

misc_feature
/note="Pfam match to entry PF00583 Acetyltransferase (GNAT) family, score 32.50, E-value 9.2e-06"
complement (3790..3813)
/gene="STY2723"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement (4420..5283)
/gene="STY2724"
complement (4420..5283)
/gene="STY2724"
complement (4420..5283)
/note="Orthologue of E. coli P76563; Faeta hit to P76563 (287 aa), 92% identity in 286 aa overlap. Contains hydrophobic, possible membrane-spanning region"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="CAD02686.1"
/db_xref="GI:16503703"
/db_xref="SPTREMBL:O824S0"
/translation="MRWQGRSNNVDEDRNRPGSPSLGGPRLPRGKGLILVVVVAAGVGVLDLGLITGQPSQOSTSISPDDEAAKFTSVIATTEEDTGGI.PQKMGKGIQKQVLRKMTTRTGCAGSWSNPFYCPADQTYIDLSFDVKKKLGADQFPAQGVIAEVEGHVQKLGIEPKVROLQONSQTEVNLSTRMELQADCFAGVGHSMQOQGVLEAGDLEALMAAQAIGDRLQOQGGRVVPSFTHTSEORYSWPKRGPDSGPAQNTFGKNP"
complement (5287..5292)
/note="possible RBS"
complement (5415..6128)
/gene="STY2725"
/note="synonym: purC"
complement (5415..6128)
/gene="STY2725"
/note="Orthologue of E. coli purC (PUR7_ECOLI); Faeta hit to PUR7_ECOLI (237 aa), 95% identity in 237 aa overlap"
/codon_start=1
/transl_table=1
/product="phosphoribosylaminoimidazole-succinocarboxamide synthase"
/protein_id="CAD02687.1"
/db_xref="GI:16503704"
/db_xref="SPTREMBL:Q8XFE9"
/translation="MQRKASLYRGKAKTVYSTENPDLLVLEFRNDTSAGDGARIKQGVIAEVEGHVQKLGIEPKVROLQONSQTEVNLSTRMELQADCFAGVGHSMQOQGVLEAGDLEALMAAQAIGDRLQOQGGRVVPSFTHTSEORYSWPKRGPDSGPAQNTFGKNP"
complement (5421..6122)
/gene="STY2725"
/note="Pfam match to entry PF01259 SAICAR synt. SAICAR synthetase, score 414.20, E-value 1.2e-120"
complement (5586..5612)
/gene="STY2725"
/note="PS01058 SAICAR synthetase signature 2"
complement (5829..5873)
/gene="STY2725"
/note="PS01057 SAICAR synthetase signature 1"
complement (6137..6141)
/note="possible RBS"
complement (6303..7337)
/gene="STY2726"

gene
Pred. No.: 2.9e-130 Length: 145050
Score: 1831.00 Matches: 365
Percent Similarity: 99.19% Conservative: 4
Best Local Similarity: 98.12% Mismatches: 3
Query Match: 98.12% Indels: 0
DB: 1 Gaps: 0

US-09-921-992-50 (1-372) x AL627275 (1-145050)

QY 1 Methisanglnalaproileglnarqlysertharqilyetvalgylasnval 20
|||||

Db 64624 ATGCATACACAGCGCTCCGATTCACAGTAGAATAATCAGATATTCCTGGGAATGCG 64565
QY 21 |||||leqlyaspqlyvalproillealvalglnsermettharsharqlytharap 40
Db 64564 CCGATTGGGAGTGGCCCCCTATCGCGTCAGTCATACATCAACGGGATCAGTGC 64505
QY 41 valglualarthaValasnlnllelysalaleugluarvalglvalaspilleaarg 60
Db 64504 GTGGAAGCAGCGCTTATCATGATCAAGCGCTGAGACGGCTTGCCACATATTTGTCG 64445
QY 61 valservalprothmetaspalalaglualapheyleuulleylsglnlnvalasn 80
Db 64444 GTTTCGTACCGACCATGACCGCCGGAAGCGTTCAACCTTATCAACGACGAGTTACG 64385
QY 81 valproleuvalalaspillehsphesaplyrargillealaleuuleuvalagluuarg 100
Db 64384 GTTCCCTGCTGTCGATATTCATCTTCGACTACCGGATTCGCTGAAGAGCGGAATAC 64325
QY 101 glvalaspCysleuargilleasnproglyasnilleglyasnlnluarqilleargmet 120
Db 64324 GCGGTGACTGCTTGCTGTTTAAACCCGCAATATCGTAAAGAACGATTCGCATG 64265
QY 121 valvalaspCysalalargasplysasnilleprolleargilleglyvalasnalaglyser 140
Db 64264 GTGGTGACTGCGCGCGCAACAAACATCCATCCGATCCGCTTAATGCGCGATCG 64205
QY 141 leuglnllyaspseuqnglnllylserlyrargllyglnprothproglmlaleuengluser 160
Db 64204 CTGGAATAAGATCTGAGAAATAATGAGAGCCACCCGAGCGCTGCTGGAATCA 64145
QY 161 Alamerarghilevalasphisleuaspargyleuasnphesaplnpheuylaserval 180
Db 64144 GCGATCGCCATGTCATCATCTCGATCTCACTTGATGATGATCAAGAGACGTA 64085
QY 181 lyealaseraspvalpheleuvalavalgluserTyrargleuvalalaysglnlleap 200
Db 64084 AAGCGCTCGATGATATTCCTTGCGCTTGATGATGATGATGATGATGATGAT 64025
QY 201 glnproleuhsileuqlylleprrglualaglyvalalargserglyalavallysser 220
Db 64024 CACCCCTGATCTGGGATACCCGAGCGGGGCGGATGACGGGGGATTAAGTCG 63965
QY 221 AlaileglyleuqlyleuuleuusergluglylleglyaspThrleuargylaserleu 240
Db 63964 GCGATCGCTTAACTGCTGCTGATGGAAGCATGCGGACACGCTGCGGATGCTTG 63905
QY 241 AlalaspProvalgluglnllelysalaglypheasplleuuleyserteuarglle 260
Db 63904 GCGGCGGACCGGATGGAAGATCAAGTCGCTTGCATATCTGAAGTCGCTGCATAT 63845
QY 261 ArgsertharglylleasnphellealacysprothCysserargnglnupheaspval 280
Db 63844 CGGCGCGCGCATCAATTCATCGCTCCGACCTGCTGTAAGAGTTCGACGTT 63785
QY 281 lleglyThrvalasnaleuqnglnarqyleuqlyaspillelleThrprometaspval 300
Db 63784 ATCGGACGCTGACCGCTGAGACGCTGCGAAGATATCATCCCGATGACGCTC 63725
QY 301 SerillelleglyCysvalvalaenglyproglyglualaleuvalaserthleuglyval 320
Db 63724 TCGATCATCGCTGCTGCTGTAACCGGCGGAGCGCTGATTCGACGCTGCGCGTG 63665
QY 321 ThrclglyAsnuleyserglyleuylrargluaspqlyvalarglyysaspargleuasp 340
Db 63664 ACGGGCGGCAATTAAGAAAGCGCTGTATGAAGACGGCTGTGAAGACGACATCGAT 63605
QY 341 Asnasaspmectilleaspqnglnuqlyualargilleargalalyalaserqnglnuasp 360
Db 63604 AACGACGATATGATCCGACGCTGATATCCGATTCGCGCGAAGATGATCAACTGAT 63545
QY 361 Glualalargargilleaspvalglnlnvalglnlyg 372
Db 63544 GAAGCGCTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 63509

Oy	361	GIJuaIaaRgCIIeAspVAlGInGlYnVaIGluYys	372
Db	14529	GAAAGCGCTGGATTGACGTCTGCAGSTTGAATAA	14494
RESULT 14			
LOCUS	AL627275/c		
DEFINITION	Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,		
ACCESSION	Complete chromosome; segment 11/20.		
VERSION	AL627275 AL513382		
KEYWORDS	AL627275.1 GI:16503698		
SOURCE			
ORGANISM	Salmonella enterica subsp. enterica serovar Typhi		
REFERENCE	Salmonella enterica subsp. enterica serovar Typhi		
AUTHORS	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella. 1 (bases 1 to 145050) Parkhill,J., Dougan,C., James,K.D., Thomson,N.R., Pickard,D., Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G., Sebatina,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T., Comerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N., Farar,J., Felwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S., Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Garra,F., Pary,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J., Stevens,K., Whitehead,S. and Barrett,B.G. Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18 Nature 413 (6858), 848-852 (2001)		
TITLE	2 (bases 1 to 145050)		
JOURNAL	Parkhill,J.		
MEDLINE	Direct Submission		
PUBMED	Submitted (25-OCT-2001) Submitted on behalf of the Salmonella		
REFERENCE	sequencing team, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK E-mail: parkhill@sanger.ac.uk		
AUTHORS	Notes:		
TITLE	Details of S. typhi sequencing at the Sanger Centre are available		
JOURNAL	on the World Wide Web. (URL, http://www.sanger.ac.uk/Projects/S_typhi/).		
COMMENT	Location/Qualifiers		
FEATURES	1..145050		
source	/organism="Salmonella enterica subsp. enterica serovar Typhi" /mol_type="genomic DNA" /strain="CT18" /db_xref="taxon:90370" 636..640 /note="possible RBS" 647..1003 /gene="STY2720" /note="synonym: yffb" 647..1003 /gene="STY2720" /note="Orthologue of E. coli yfjB (YFPB_ECOLI), Fasta hit to YFPB_ECOLI (118 aa), 81% identity in 118 aa overlap" /codon_start=1 /transl_table=1 /product="conserved hypothetical protein" /protein_id="CAD02682.1" /db_xref="GI:16503698" /db_xref="SPTRMBL:O82453" /translation="MTITGLIKNCDDTKKARMLDEHGIDYRFHYRVVDGIDLPLNT FLIALGWQPLINTGRGTWRKLDEBHRSGITDADDAALMLVGPRIIRPLILCAKGKM LLGESRSRYQQFDEV" 995..1000 /gene="STY2720" /note="possible RBS" 1007..2134 /gene="STY2721" /note="synonym: dape"		
CDS	1007..2134		
RBS			
gene			
CDS			

misc_feature

/gene="STY721"
/note="Orthologue of E. coli daps (DAP_ECOLI); Pasta hit to DAP_ECOLI (375 aa), 94% identity in 375 aa overlap"
/codon_start=1
/transl_table=1
/product="succinyl-diaminopimelate desuccinylase"
/protein_id="CAD02683.1"
/db_xref="GI:16503700"
/db_xref="SPTREMBL:Q8Z4S2"
/transl_table=1
/note="MSCPVILITQOLIRBPSLDVDGCOALMIERLRKIGTIEHND FGDTPNFAMAGRGETTLAFAGHTVDVADWDVINPPEPTTRDGWLPRGADMGV SLAAWVAAEFPAOHPIHRRLAFLITSDSEASAKSGTKAVYEAALKARBLDYCVG GEPSTETPGVVNKKGRGSITCNLTTHGOGHYAVPLADNPHPRAPLINLEYALE WDRGDFPFATPSMOTAINOAGTSNNMTIPGBLPFYNPNRSTETIDEMIKERHALLE KHQLVTYDDWMISQGPLTAGRKLVDAVNVAIEHVNSIKPOLITGTSGDRFLARG AQVELGPVNATIHKINECVNAADLOLLARMYORIMEQLVA"
1031..1978

misc_feature

/gene="STY721"
/note="Pfam match to entry PF01546 Peptidase_M20, peptidase family M20/M25/M40, score 315.30, E-value 7.4e-91"
1187..1216

misc_feature

/gene="STY721"
/note="PS00758 Arge / dape / ACY1 / CPG2 / yscs family signature 1"
signature 1"
1295..1414

RBS

/gene="STY721"
/note="PS00759 Arge / dape / ACY1 / CPG2 / yscs family signature 2"
signature 2"
2150..2153

BRS

/note="possible RBS"
2162..2362

gene

/gene="STY722"
2162..2362

CDS

/gene="STY722"
/note="similar to Escherichia coli similar to TR:p77259 [EMB1:D50875] (66 aa) fasta scores: E(): 1e-27, 95.5% id in 66 aa. Contains hydrophobic, possible membrane-spanning region"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="CAD02684.1"
/db_xref="GI:16503701"
/db_xref="SPTREMBL:Q8XP02"
/translation="MDMLAKMWILTVLVLGVLLIVKDLKRDHKFKPLANKELPP HRDNFDKDDEDDMPKKDOPKK"
complement(2389..4404)
/gene="STY723"
complement(2389..4404)

gene

/gene="STY723"
/note="Orthologue of E. coli YPFI_ECOLI; Pasta hit to YPFI_ECOLI (671 aa), 72% identity in 662 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD02685.1"
/db_xref="GI:16503702"
/db_xref="SPTREMBL:Q8Z4S1"
/translation="MSDIDRALITSOMTOEGRIILLVISGDAMCGERAEAIRAVLP GDWWAPAPAPOPRCTPOALQTLLGREFRHAIFDWGCFDPAAPALASGLQAGSWLL LLMPEYWEESRPDTSLRWSDCAQPIPTPOFOQHAKRTISRDPOTLLMRPORPCWP SPYSERRPATGPEPOEOAAILSRLEMPEPVATVAPARGSKSLAAGOFISOMGT ALIVAPADITDIILAFAGGERFCFMADALLASGARADMVVDVEAAIPPLILOVS RPPELLTTTYQGYEGTGRCFLKFCARPQIARFTLRQVRNAPCEPLENTISEALI FDDEFAPAPGGIAISAFTQQANGKTPALPRAVYTQLSSAHRTSPSLRRMKWDVG QHFLQATNMNVAGALWVEBGISAELSQAWCGRFPFGNIWAOSLAHGSDPLAA TLVGRYSRIIVHPAROREGIGQOLICACEOAQCCTSVSGFTPELARFWORGCF VLVNRGSHRESGCYTMALPLPSDAGSLAOOEHRRLRADDIITONNGEAIPIAA LREAADNDEMRRELTCFAFAHRPLITSGCHRLTLYSALPLPALGRLEEKSNDEL CARLAISGRKALLALOAPAQAQALITALDAORTLRDWNCGGDHAG"
complement(2878..3207)
/gene="STY723"

misc_feature

/db xref="GI:16421061"
/translation="MVRRIILVILVSVVAAGDNGSANDSHSSLPDLASES
AKKEBENKSKREKAGDYFINSATGEPNLTPEALREABRYLQNISSASQYLE
GMSPYKRTITSLSVEGDLDSSTLFTIPWTDNSTLFSQISAKCKEDRTINGRG
LGVQVQGNVLGLGNAYDYDFRGHRLGLGEAMVDYKFSGNTYHPLSDWKEDE
FDEYEREPARGMDIIRMSMLPEYQOLGAKLYEQYDGEVALGTDMLOQDPDAVL
LEVTPLVLTVTGTDYKAGTSDNSDFSNATVANYOIGPLAOLDEPVKIOHLSMR
TDFVDNNFTILEYREKDPDLVTLMKADATNHPCEVIDTPEAALGLEKCKMTYNA
LIMHYKTISSAQKNNAKNAITLMPVYKXNALTEGNSGWNLTLPAMVADPREORT
ALNTWKRTMTLEDEKKNQNSGVETITVOODRKIELVDNIADTDSDSHESALAD
GEGVAVDILITDSFGDSTDRNGNELVDAMTPELVYDSNCKVTLAQTCTTPTPCV
IASRDEKAGVTLLSTLPGTFRMKAKADAGDSNYDVDTIGNLSALNVAIOVAA
NPVNLGKEDKHPTVNNAYRFLMRDNKGVGOMSPQLSEEMALDYDMWFTGOST
NGHTGALANTMEDLVLPVTNKEAOKFAANVEDGVGVGIVRTYSGK"
/complement (9037..9042)
/gene="sinh"
/note="putative RBS for sinh; RegulonDB:STWS1H002624"
/complement (9382..9615)
/gene="yfgJ"
/note="synonym: STM2518"
/complement (9382..9603)
/gene="yfgJ"
/note="similar to E. coli orf, hypothetical protein
(AAC75563.1); Blastp hit to AAC75563.1 (83 aa), 63%
identity in aa 13 - 83"
/codon_start=1
/trans_table=1
/product="putative cytoplasmic protein"
/protein_id="AAL21412.1"
/db xref="GI:16421062"
/translation="MEITCPVCHHALERNQDTAHCTCAKDFLQALCPDRCPLQVL
KACGAVDYFCONGHGLISKRVNPLSDQ"
/complement (9610..9615)
/gene="yfgJ"
/note="putative RBS for yfgJ; RegulonDB:STWS1H002625"
/complement (9692..11176)
/gene="engA"
/note="synonym: STM2519"
/complement (9692..11164)
/gene="engA"
/note="similar to E. coli putative GTP-binding factor
(AAC75564.1); Blastp hit to AAC75564.1 (503 aa), 97%
identity in aa 14 - 503"
/codon_start=1
/trans_table=1
/product="putative GTP-binding protein"
/protein_id="AAL21413.1"
/db xref="GI:16421063"
/translation="MVPVVALVGRPNVSKSTLFRRLTRTRDALVADPPGLTRPKKGR
AYVGRFFCTIDGIDGTEDRGVETMAEBSILAIEADVVLFMVARAGIMADEAI
AKHLRSREKPTPLVANKTDLDPQAVVDYSGIGITVPIAASHRGVLSLEHVL
PMWDVAPQSEVEDADAYMAQFEAQGESEAPEDDPQSLPKLALVGFNPKSTL
TNRLGERRVVVDMPTGRDSIYIPMERDEREYVLDITGVRRKKITDAVERFSV
KTIQALIEDANVLLVIDAREGISDQSLGFIINSRSLIVVNKKDGLSQVEKEQV
KETIDFLRGIDFARVHFISALHSGVGNLFESVREAYDSTSRVSTAMLTRIMTAV
EDHOPPLVGRGRVKKYAHAGVNPPIVHGNQVLDLPSTYRXYLMNVRKSLFWMG
TPRIQPKESENPRANKRNTLTPQMKRKLKMKIKKSK"
/complement (11171..11176)
/gene="engA"
/note="putative RBS for engA; RegulonDB:STWS1H002626"
/complement (11283..12471)
/gene="yfgL"
/note="synonym: STM2520"

Alignment Scores:
Pred. No.: 5.3e-132 Length: 23647
Score: 1840.00 Matches: 367
Percent Similarity: 99.46% Conservative: 3
Best Local Similarity: 98.66% Mismatches: 2
Query Match: 98.61% Indels: 0
Gaps: 0

US-09-921-992-50 (1-372) X AE008814 (1-23647)

QY 1 MetHisAsnGluAlaProIleGlnArgLysSerThrArgIleTyrValGluAsnVal 20
Db 15609 ATGCAATACAGAGCTCCGATTCAACGTAGAAAATCAACAGTATTACCTGGGAATGTC 15550
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 15549 CCGATTGGCGATGGTCCCTTATCCGCGTCAATGATGATCAACGCGTACCATGAC 15490
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGluAlaAspIleValArg 60
Db 15489 GTGGAAAGCCACGGTTATATAGATCAAAAGCGCGAGACGGCGTTGGCCAGATATTGTCGT 15430
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
Db 15429 GTTTCGTACCAACATAGAACCGCCCGCAAGCCTTACAGCTTACAAACGACGGTTAAC 15370
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db 15369 GTCCCGCTGGTGGCGATATTCACTTCGATACCGCATTGGCGTGAAGTAGCGAATAC 15310
QY 101 GlyValAlaAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluArgIleArgMet 120
Db 15309 GCGGTGACTGCTTCGCTATCAACCCCGCAATATCGTAAACGAAAGCTATTCTGATAG 15250
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
Db 15249 GTGTGACTGCGCGCGCAAAACATCTTATTCGATCGCGCTAAACCGCGATCG 15190
QY 141 LeuGluLysAspLeuGlnLysTyrGlyGluProThrProGluAlaLeuLeuGlnSer 160
Db 15189 CTGAAATAAGATCTGACGAAATAATCGGTAGACGACCGCGACGCGTGTGTAATCA 15130
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
Db 15129 GGCATCGGCATGTCATCATCTCAATCGCTCACTTGTATCAAGTCAAGACGCTA 15070
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 15069 AAGAGCTCCGATGATATCTCTGCGCTTGAGTCTTATCGTTGTGGCTAAACGATCGAT 15010
QY 201 GlnProLeuHisLeuGlnYlIethrGluAlaGlyValArgSerGlyValAlaLysSer 220
Db 15009 CACCGCTCATCTGGGATCAACCGAAGCGGGCGCGGTAGCGGGCGGTCAAGTGC 14950
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
Db 14949 GCGATCGGTTTAGTCTGCTGTTATCGGAAGGATCGGAGACACGCTGCGCGTATCGTTG 14890
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 14889 GCGGCGGACCGGTGGAAGAGATCAAAAGTGGTTGATGATTTGAAAGTCCGCGCATTT 14830
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnLysPheAspVal 280
Db 14829 CCGGCGCGGCGGATCAATTTCATCGCTCCGACCTGTTCTCGCAGAGGTTGACGATT 14770
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
Db 14769 ATCGGACCGGTGAACCGCTGCGAGCGCTCGGAAGATATCATCAACCGAGAGACGTC 14710
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGluVal 320
Db 14709 TCGATCATCGGTTGCTGCTAGTAACCGGCGCGGCGGAGCGCTGATTCAGCGTGGCGGTG 14650
QY 321 ThrGlyValAsnLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
Db 14649 ACGGCGGCAATATAGAAAAGCGGCTGTATGAAGACGGGTGGTAAAGACAGACTCGAT 14590
QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaValAlaSerGlnLeuAsp 360
Db 14589 AACGACGATATGATCGCGACGCTTGAATCCCGTATCGCGCGAAACAGTCACTTGAT 14530

Dn	189 AACACGATTCAGTGCACCTGGAAACACGACTTCGTCCGAAGACCAGTCACTGAC	130
Oy	361 GUAAGAAGAGLLEAPVAILGNGINVAJGLUYLS	372
Dn	129 GAACGCGCTCGAATTGACGTTCAGACAGGTGAAAA	94
RESULT 13		
AEO08814/c		
LOCUS	AEO08814	23647 bp DNA linear BCT 23-APR-2003
DEFINITION	Salmonella typhimurium LT2, section 118 of 220 of the complete genome.	
ACCESSION	AEO08814	AE006468
VERSION	AEO08814.1	GI:16421058
KEYWORDS		
SOURCE ORGANISM	Salmonella typhimurium LT2 Salmonella typhimurium LT2 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.	
REFERENCE AUTHORS	1 (bases 1 to 23647) McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W., Latreille,P., Courtney,L., Potwollik,S., Ali,U., Dante,M., Du,F., Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A., Grewal,N., Milvanev,E., Ryan,E., Sun,H., Flores,L., Miller,W., Stoneling,T., Nhan,M., Waterson,R. and Wilson,R.K. Complete genome sequence of Salmonella enterica serovar Typhimurium LT2	
TITLE	Nature 413 (6858), 852-856 (2001)	
JOURNAL MEDLINE	21534948	
PUBMED	11677609	
REFERENCE	2 (bases 1 to 23647)	
AUTHORS	The Salmonella typhimurium Genome Sequencing Project	
CONSRM TITLE	Direct Submission	
JOURNAL	Submitted (29-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108 USA	
COMMENT	Supported by NIH grant SU 01 AI43283	

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GUMMER; <http://www.tigr.org/software/glimmer/glimmer.html> and GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and Regunold;

http://kinich.cifn.unam.mx:8950/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

Location/Qualifiers

```

1..23647
/oranism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/restrain="LT2: SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
complement(110..5711)
/gene="rfaC"
/note="synonym: STM2515"
complement(110..5707)
/gene="rfaC"
```

FEATURES

Source

gene

CDS

/note="Rat4 [gi|4583529] "
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAL21409.1"
 /db_xref="GI:16421059"
 /translation="MDRRROYKVYHFLTLTIMFAVSIAPQALLKGGTMOELNSVTL
 GAVNTATPLAAGAITPLYGQSTLIDPSKTHIDIESAMPDRPSDASATSTSKRAVNSTDTDD
 EGDLESDPEPTIAMENRQEPAMGLWADADITPLASPOGFENLFLFCQNLGVOLVAM
 AQVEDEIVPALMEFTRTGVNVAITPLSLKVAIIRKPAIDPVTVYSGHNVDAFEPA
 SKVQGESITLTIITKACDGTPEGNGAFVIRKDAENRQGVNNANFVRVDDTLTTT
 CTREYGVADAGNATVVVTOKEGNGVTPVLVSVSNPALFAETAVIETITTSDDSP
 ASMGMHIESATITLNGITETFTFTRPKLAESGADKSVDPNEMTALPHMSAGNHDD
 ILPDARQIVQKREHSTLETKTGMVAGDAEYMSSTKQSLGTHAAVHMSASVYVRRA
 DSDTLVSCVQKRAHPQITLSPQGYKQVSESILDTMTVDROTQKPLPYRME
 LEPLPAKRKKEGHDAMDNLNRYTVDSEMSASSEBHYTGVDVNGOAILTKHNSGLG
 VETPIRPMPEDEGENVELSPSVITFTVTSDDVGGAMMGHMRGVDAAGILYKPLILA
 VEAHSKQDSQSENNEMWATFNSVASATQCGVGVPOOSLAHLYSEHAGGQMSHEG
 MPTBEYPIAADSSDAGTHVLENGDSGKPTDPNTYILTCSANEMVALIDYFENDPPI
 KNADKVEQYKRNMIHSHRNALNGAIGINTDPTTMANGRRKRDGLTGETDTSNCEMP
 KADKVEQYKRNMIHSHRNALNGAIGINTDPTTMANGRRKRDGLTGETDTSNCEMP
 DDVGVAAQVYHGTTDANGATITLTQKQVGLTTPNLIAVDSLSTPSRSYKFTV
 AISPPTPAMPMGMHADITVGDWTFEPKPLAGSNMPLRQDSNETMVAHSDVAV
 GNPDGCGAARLPRIDQLBALYNVNASGKILHSGMPTVSYOSATTKLJAL
 LTNSEFASGVNLSIYASCLASDNVVASITHEPYNQWNGSVYHVAKKGTETWLO
 KYTVKDSGNIPEAPFVLTGRGDCYDRGEXYTKQDDDLQGITPVIVIGESLAWTT
 TKMSQSTDTGRTIIVTRPDTHGRTALNITLLENAVSASIDITFTVTSPPVSAVA
 RMWGMASLITADGAVYQRLPLAELSTDNVTSKQETMETAVFPGALDEBGNAPAR
 CAAGYPVAEALDSLVSYSKYPISRTINTAQGMFVYVYSGNSITSFSSGAKLDEYAAD
 LADGSRSENETSTAMQVOICATTPLEQATQITLTSPOAMDMDIAQVKAHNSIPL
 LITTTDAGNVPVATPSLRKDAKANPDNKFVALNGTNMTVTLTGAAQOCQVYVAT
 SVLTGATGADDTLALTLAEPGICLKNOUTLNPATPSNLSPVFTVLTSPSDAKA
 NMGYMPTFPASGAETKRLPVAGEPSESEHTDITYETENEMIMVNSFNTGNGGCP
 NMQMAIDDFALYNDRHSRGVATDIGLPVAKRWADSLSKGSTLTYQKDLTKGNK
 YMSNPNENYVLOCLTTPSRSGNIALTSDSMANDKSAVAAKKSETIPMTYVINDGK
 QOPQALVILTRDYASGAVDKQYIEPGVGEVSPPTTSPANMLAPVAPAGVAVAV
 NNQNGSLSTKMSGSGVDGCKLFTLTQKSGIKGLTSVTAALANQMDASVADITVOTD
 SPDIPTVSYGMHMDYQVNGVTILRPYLAKELSAARDPMENFENFEGNYYVQSHSH
 VETSIPTLCSGQENIASLDDLKALQSVIQTILQMPFTTSSMDIVSDQESKNKYCSFNE
 TTQTQTTREKATTSGLGSCRP"
 complement (5706. .5711)
 /gene="rfaA"
 /note="putative RBS for rfaC; RegulonDB:STMS1H002622"
 complement (5827. .6793)
 /gene="sin1"
 /note="synonym: STM2516"
 complement (5827. .6786)
 /gene="sin1"
 /note="sin1 [gi|4583530] "
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAL21410.1"
 /db_xref="GI:16421060"
 /translation="WQATVYKRLTKVALALVAGYCAAPVAANGNLSKSGOMQVSEQ
 TGIYQGVIPNATRAADTKADDKDHTVITIDRGKRLVTEDQCFHGVGATVYMAIG
 DTEGDLDTNDATRAATYKWSFSDQNSVSDDLCTGQSYEIQADADARYIGIKITPTT
 TTGPATVATLILKDLSTGAGSGSDDEIEGPGVDENVVVIHEKDSNNLKNSSCT
 TLKNTITVOVLLMLKDGNGKNTYDAGENVQDYRMKFGVSKLAGTGGIIVNESVN
 DKDLIVPTNNAETAFEGADGVTGVSDDQSGSLSDYKRR"
 complement (6788. .6793)
 /gene="sin1"
 /note="putative RBS for sin1; RegulonDB:STMS1H002623"
 complement (6844. .9042)
 /gene="sinH"
 /note="synonym: STM2517"
 complement (6844. .9036)
 /gene="sinH"
 /note="similar to Escherichia coli incuinin and Yersinia
 pestis invasins proteins; [gi|4583531] "
 /codon_start=1
 /transl_table=11
 /product="sinH"
 /protein_id="AAL21411.1"

/transl_table=11
/product="ISSf14 orf"
/protein_id="AAPI7892.1"
/db_xref="GI:30042167"
/translation="MTESDSYVESVQVFIGVDVKDTHHAAVAINRSGKRLFDKALPNDE
NKRSLISDKQHGQILVVDOPATIGALVAARSGVAVGLPGIAMRRIADLHAG
EAKTDVDAAIIEAARTLPHALRTLDQEAIAELMCGPDDDLAAQTOQASNRIR
GLATQIHPAPERIVGPRLEHVAVIDLQRYSPESPKLASGEKLAOLCGIAPRLGKR
LAADIQALAEOTVVVGTGNAAVVLPRLALCLITLRKQDEVALAEQVAVLHXP
VUTSMRGVGYRTAARLITVACRAFASAAHLAAYAGIAPTRRSGSIREHPSRRGN
KAKRAFLFSAFALNDPLSRAYYTRKMSQGRHNOALALARRCDVLFAMWRDGLF
YTPQGS"
7960..8442
/gene="ssea"
/locus_tag="S2740"
/note="truncated"
/pseudo
complement (9261..10037)
/gene="sseb"
/locus_tag="S2741"
complement (9261..10037)
/gene="sseb"
/locus_tag="S2741"
/function="phenotype; Not classified"
/note="residues 1 to 258 of 258 are 94.96 pct identical to
residues 4 to 261 of 261 from Escherichia coli K-12 :
B2522"
/codon_start=1
/transl_table=11
/product="enhancer of serine sensitivity"
/protein_id="AAPI7893.1"
/db_xref="GI:30042168"
/translation="MSETKNELDLLEKATEPAPRPAFRTLESTVWVPGTAQGE
AVVEDBALDIQHEKEDGTSVITPFTSLLELQAVDEQAFVMPRTTLEMTLGETL
FLNAKLPTEGEMPRETISLTIGEENPLSQETLEGGESLISVAEPDPMQDSLT
LFTTIRKVKAFICSIKEMEAQPNLIGLEADGIEIIOAGSVATDTLPDDEID
ICQKKGKGISHTFTEHIAFERRGGCFRDFKQNRIT"
10137..10170
/note="RBP (repetitive extragenic palindromic) element,
contains 1 RBP sequence"
complement (10179..11462)
/gene="pepb"
/locus_tag="S2742"
complement (10179..11462)
/gene="pepb"
/locus_tag="S2742"
/function="putative enzyme; Not classified"
/note="residues 1 to 427 of 427 are 92.50 pct identical to
residues 30 to 456 of 456 from Escherichia coli K-12 :
B2523"
/codon_start=1
/transl_table=11
/product="putative peptidase"
/protein_id="AAPI7894.1"
/db_xref="GI:30042169"
/translation="MTAEAMKITLSTOPADARWGEKATYSINNDGITLHNGADLGI
ORAAKIDLCIGIKVQLSGGMDADKRWANQCYKPKGRKREMDIDDAQCELDN
RLMTIDVDRITNAPBELSPQLAQAVDLISNAGDRVTYRITKEBDRREGYGL
HTVGRGSESPVLLADLYNPTGDKAAYVCLVKGITPDSGGYSIKQTAFMDSKMSD
MGGAATVGTGALAFIMRGLNKRVLFLCCADNLISGNAFLGDIITRYNKKYEVNAT
DARGSLVADGLIDASKQPEMTIDAATLTGAATKALGNDYHALFSDDDALARLLAS
ASQENEFMRLLPLAEFHSQLPSPNPAELNNTGSAAYPAGASTAAGFLSHVENYQCGM
LHIDCATYRKAVPEQMSAGATGLGRTINLNLTA"
complement (11521..11721)
/gene="yfhj"
/locus_tag="S2743"
/gene="yfhj"

Alignment Scores:
Pred. No.: 1e-131 Length: 290380
Percent: 1855.00 Matches: 370
Score: 99.73% Conservative: 1
Best Local Similarity: 99.46% Mismatches: 0
Query Match: 99.41% Indels: 0

DB: 1 Gaps: 0
US-09-921-992-50 (1-372) x AEO16987 (1-290380)
Qy 1 MethiAsnGlnAlaProIleGlnArgLysSerThrArgIleTyValAlaGlnVal 20
Db 1209 ATGCATTAACAGAGCTCCATTCAACGTAGAAATCAACACGTATTACGTGGGAATGTG 1150
Qy 21 ProIleGlyAspGlyValaProIleAlaValaGlnSerMetThrAsnThrArgTrpTrp 40
Db 1149 CCGATTGGCGATGGTGTCTCCATCGCGGTACGTATCATACCAATACGCTAGACAGAT 1090
Qy 41 ValGluAlaThrValaAsnGlnIleLysAlaLeuGluArgValaGlyAlaAspIleValArg 60
Db 1089 GTCCAGCAACGCTCATCAATCAAGCGCTGGAACGGCTTGCCCTGATATCGTCCG 1030
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValaAsn 80
Db 1029 GTCTCCGTACCGACGATGACGCGGAGAACGCTTCAAACTCATCAAAACGCGGTAAAC 970
Qy 81 ValProLeuValAlaAspIleHisPheAspTyArgIleAlaLeuLysValaAlaGluTy 100
Db 969 GTCCGCTGCTGCTGACATCACTTCGACTATCGCATTCGCTGAAGTAGCGGAATAC 910
Qy 101 GlyValaAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
Db 909 GCGTCGATTGCTGCTGATTAACTTGGTATATCGTATATCAAGAGCGTATTCGAT 850
Qy 121 ValValaAspCysAlaArgAspLysAsnIleProIleArgIleGlyValaAsnAlaGlySer 140
Db 849 GGTGTTGACTGCGCGCGCATAAACATTCGATCGTATTGGCTTAACGCGGATCG 790
Qy 141 LeuGluLysAspLeuGlnGluLysTyrgLysGluProThrProGlnAlaLeuLeuGlnSer 160
Db 789 CTGGAATAAAGATCTGCAATAAAGATGAGCAACGACCGCGAGCGTGTGCGAATCC 730
Qy 161 AlaMetArgHisValaAspHisLeuAspArgLeuAsnPheAspGlnPheLysValaSerVal 180
Db 729 GCATCGCTCATGTTGATCATCTCGATCCGCTGAATTCGATCAGTTCAAAGCAACGCTG 670
Qy 181 LysAlaSerAspValaPheLeuAlaValaGlnSerTyArgLeuLeuAlaLysGlnIleAsp 200
Db 669 AAAGCGCTGACGCTCTCCCTGCTGTGAGTCTTATCGTTGCTGCAAAACAGATCGAT 610
Qy 201 GlnProLeuHisLeuGlyLysLeuThrGluAlaGlyValaArgSerLysValaLysSer 220
Db 609 CAGCCCTCATCTGGGATCACCGAAGCGGTGTCACGACGCGGCGACGTAATAATCC 550
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValaSerLeu 240
Db 549 GCCATTGTTAGGTGCTGCTGCTGCGAAGGATCGCGACACGCTGCTGATCGCTG 490
Qy 241 AlaAlaAspProValaGluGluLysValaGlyPheAspIleLeuLysSerLeuArgIle 260
Db 489 GCGGCGATCCGCTGCAAGATCAAGATCGGTTGATTTGAAATCGCTGCGCTATC 430
Qy 261 ArgSerArgGlyLysAsnHeterLeaLysProThrCysSerArgGlnGluPheAspVal 280
Db 429 CGTTCCGAGGAGTTAATCTTCACTGCTGCCACCTGTTCCGTGAGGAATTTGATGT 370
Qy 281 IleGlyThrValaAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
Db 369 ATCGGATCGGTTACCGCTGAGCAACCGCTGGAAGATATCATCTCCGATGAGACGTT 310
Qy 301 SerIleIleGlyCysValaValaAsnGlyProGlyGluAlaLeuValaSerThrLeuGlyVal 320
Db 309 TCGATTATCGGCTGCGCTGTAATGAGCCAGGAGAGCGCTGTTCTACACTCGCGCTC 250
Qy 321 ThrGlyGlyAsnLysSerGlyLeuTyrgLysGlyValaArgLysAspArgLeuAsp 340
Db 249 ACCGGGGGCAACAGAAAGCGGCTCTATGAAGATGGGTGGCAAAACCGCTGAGAC 190
Qy 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360


```
/product="orf, conserved hypothetical protein"
/protein_id="AAN4059.1"
/db_xref="GI:24052933"
/translation="MEIYENENDQVEAVKFPFANGKALAVGILGVALTGRYKNS
HVDVASASALAYQNAVAVSEKPDSPFAEKFAENKTYGALSLELAQDFVKX
ELERKAQLOQGLADSDENLKAVINILARVVOLOKADAPLKTIDTITIGREMAAIV
ADIRGEALISKSGKOGARSAMEAGVSDVTPALSEMMQMKINILSI"
complement(5138..6412)
/gene="hies"
/locus_tag="SF2560"
complement(5138..6412)
/gene="hies"
/locus_tag="SF2560"
complement(5138..6412)
/notes="Residues 1 to 424 of 424 are 99 pct identical to
residues 1 to 424 of a 424 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_289067.1"
/codon_start=1
/transl_table=1
/product="histidine tRNA synthetase"
/protein_id="AAN4060.1"
/db_xref="GI:24052934"
/translation="MAKNQATIRGANDYLPGETAIWRIETGLKAVLGSGYEILP
IVEQTPLEKAIQEVLDVVEKEMYTEEDRGDSLTLRPECTACVRAGLEHLLYQE
ORLMTYIGPMFRHERPOKGRYRPHQEGEVGLOGPDIDAEILMLTARMRALGIEH
VTELEINSIGLEKARANYRDAVAFLEQHKLEDECKRMYTNPRLVLDKMEVVAL
LNDAPALGDYLDSESRHFAGLCKLESAGIAYTVNRIYRGIDYVNRATKPEVNTSL
GSGCTVAGGRPGVGLVYEQLEGRATPANGMGLPRYLIVQANPERKADPDYILV
ASGADTOSAMALAEIRLDELGVKLTNHNHGGNFKKQFARADKMGARAVAVLGESEV
ANGTAVVKDLRSEQITVAQDSVAAMLRLTLG"
complement(6515..7642)
/gene="gcpB"
/locus_tag="SF2561"
complement(6515..7642)
/gene="gcpB"
/locus_tag="SF2561"
/notes="Residues 1 to 372 of 375 are 99 pct identical to
residues 1 to 372 of a 372 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_289068.1"
/codon_start=1
/transl_table=1
/product="orf, conserved hypothetical protein"
/protein_id="AAN4061.1"
/db_xref="GI:24052935"
/translation="MHNQARIORRKSRIIVGNVPIGDAPIAVOSWTRTDEAT
VNOIKALERGADIVRSVPTMDAAEFKIKORAVNPVADIHEDRIALKABEYGV
DCRIINPENGINERIRIMVVDCAKDKITPIRIGVNSSEKEDQEKIGEPPLAES
AMRHVHLDRLENDPQFVSVKASDVLAESYRLAKQIDQPLHGLTTEGASGAV
KSAIGLILSEBIDGLRVLASADPVEIKVGDILKISIRSGINFTACPTCSRO
EFVIGTVNALBOLEBDITPMDSIIGCVNPGSEALVSTLIGVTKGKKSGLYEDGV
RKORLDNNMDITDLEARIKAKASQLEARRIDVQVGIIT"
complement(7669..8682)
/gene="yfgA"
/locus_tag="SF2562"
complement(7669..8682)
/gene="yfgA"
/locus_tag="SF2562"
complement(7669..8682)
/notes="Residues 1 to 337 of 337 are 99 pct identical to
residues 1 to 337 of a 337 aa protein from Escherichia
coli O157:H7 ref: NP_311405.1"
/codon_start=1
/transl_table=1
/product="putative membrane protein"
/protein_id="AAN4062.1"
/db_xref="GI:24052936"
/translation="MNTTEATHDONEALTGARRLNAREQGLSQQAVAERLCUKVSTV
RDIEDKAPADLASTPLRGYIRSYARLVAIPBEELPLGELKQAPLRAPKAPQGSFL
GKRKRKDGWIMETWLVLFVYIGLSGAMMMDHKAQOEIITMAADSAELESNSSE
GQSVPLNTSTTDPATSTPASYVDVATNTATOTPAVAPAPADVDPOONAVASQONV
DTNATPVPATATTPDGAAPLPDQAGVTTAADBNALVMTFADCNLEVDATGKCLF
```

Alignment Scores:
Pred. No.:
Score:

1.2e-133 Length: 10225
1855.00 Matches: 370

```
Percent Similarity: 99.73% Conservative: 1  
Best Local Similarity: 99.46% Mismatches: 0  
Query Match: 99.41% Indels: 0  
DB: 1 Gaps: 0  
US-09-921-992-50 (1-372) x AB015271 (1-10225)  
Qy 1 MeHh:ASnGLAAlaProIleGlnARGLYSerThrArgIleThrValGlyAenVal 20  
Db 7642 ATGCATACCGAGCTCCATTCATTCAGGTAAGAAATCAACGATTTACCTGGGATATG 7583  
Qy 21 ProIleGlyASPGLYAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40  
Db 7582 CGGATGGGAGATGGTCTCCATCCGCGTACATGATCAACCAATCGGTAACGACAGAT 7523  
Qy 41 ValGluAlaThrValAlaSerGlnIleLysAlaLeuGluArgValGlyAlaSerIleValArg 60  
Db 7522 GTCGAAGCAACGGTCATCAATCAAGGCGCTGGAAAGCGTTGGCCCTCATATCGTCGGT 7463  
Qy 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnIleValAan 80  
Db 7462 GTCTCCGTAACCGATGAGATGAGCGGCGAAGAGCTTCAACTCATCAACGCGGTTTAC 7403  
Qy 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100  
Db 7402 GTGCGCGTGGTGGCTGACATCCACTTCGATTCGATTGCCGTGAAGTAGCGAATAC 7243  
Qy 101 GlyValASPcysLeuArgIleAsnProGlyAsnIleGlyAsnGluLysIleArgMet 120  
Db 7342 GCGGTGATGTCGTGGTATTAACCTCGTAAATATCGTAATGAAAGGCGTATTCGCATG 7283  
Qy 121 ValValASPcysAlaArgAspLysAsnIleProIleArgIleGlyValAsnIleGlySer 140  
Db 7282 GTGTTGACTGTGGCGGCAATTAACATTCGATCGATTTGGCGTTAACCGCGGATCG 7223  
Qy 141 LeuGluLysASPleuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160  
Db 7222 CTGGAAGAAAGATCTGCAAGAAAGATATGCGAAGACCGCGGCGTTTCGCGAATCC 7163  
Qy 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180  
Db 7162 GCATCGCATGATGTCATCATCTCGATCCCTGAACTTCGATCAAGTCAACGACCGTG 7103  
Qy 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200  
Db 7102 AAGGCGTCTGACGCTTCTCTCGTGTGAGTCTTATCGCTTTCGCAAAACGATCGAT 7043  
Qy 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyValAlaLysSer 220  
Db 7042 CAGCGCTCATCTGGGATCAACCGAAGCGGTGTCACGACGCGGCGGCAATTAATCC 6983  
Qy 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240  
Db 6982 GCCATTGCTTATAGTCTGCTGCTGCGAAGGATCGGCGACACGCTGCTGATGCTG 6923  
Qy 241 AlaAlaAspProValGluGlnIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260  
Db 6922 GCGGCGGATCCGTCGCAAGGATCAAGATCGGTTTGATTTGAAATGCGTTCGATTC 6863  
Qy 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGluPheAspVal 280  
Db 6862 CGTTGCGAGGAGATTAACTTCACTCGCTCCGCACTGTTCCGTCAGGAATTGATGTT 6803  
Qy 281 IleGlyThrValAsnAlaLeuGlnGluAlaGluLeuAspIleIleThrProMetAspVal 300  
Db 6802 ATCGGATCGGTTAACCGCTGCGAAGACCGCTGGAAGATATCATCTCGAAGGACGTT 6743  
Qy 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320  
Db 6742 TCGATTATCGGCTGCGGTGATGAGCCGAGGAGGCGCTGTTCTACACTCGGCGTC 6683  
Qy 321 ThrGlyIleAsnLysLysSerGlyLeuTyrGluAspGlyValArgLysAspArgLeuAsp 340
```

QY 321 ThrGlyAlaAsnYsLysSerGlyLeuTyGlnuAspGlyValAlaGlyAspArgLeuasp 340
Db 187450 ACTGGCGCAACAAGAAAGCGGCTCTATGAGATGCGCGCCCAAGACCGCTTGAC 187391
QY 341 AsnAsnAspMetIleAspGlnLeuGlnuAlaArgIleArgAlaValAspGlnLeuasp 360
Db 187390 AACCAACATATGATGATCGATCGCTGGAGCGCGCATTTGTCGGAAGCCAGCTTGAC 187331
QY 361 GlnuAlaArgArgIleAspValGlnGlnuAlaGlnuAsp 372
Db 187330 GAAGCGCTGCAATTGACGTTCCAGCAGCTTGAAAA 187295
RESULT 11
AE015271/c
LOCUS
DEFINITION
Shigella flexneri 2a str. 301 section 234 of 412 of the complete
genome.
ACCESSION
AE015271 AE005674
VERSION
AE015271.1 GI:24052928
SOURCE
ORGANISM
Shigella flexneri 2a str. 301
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
REFERENCE
AUTHORS
Jin, Q., Yuan, Z. H., Xu, J. G., Wang, Y., Shen, Y., Lu, W. C., Wang, J. H.,
Wu, H. T., Dong, J., Sun, L. L., Xue, Y., Zhao, X. B., Zhang, J. Y., Yang, G. W.,
Kan, B., Chen, S. X., Yao, Z. J., He, B. K., Chen, R. S., Ma, D. L.,
Qiang, B. Q., Wen, Y. M., Hou, Y. D. and Yu, J.
Genome sequence of Shigella flexneri 2a: insights into
pathogenicity through comparison with genomes of Escherichia coli
K12 and O157
Nucleic Acids Res. 30 (20), 4432-4441 (2002)
JOURNAL
PUBMED
2384590
REFERENCE
AUTHORS
Jin, Q., Shen, Y., Wang, J. H., Liu, H., Yang, J., Yang, F., Zhang, X. B.,
Zhang, J. Y., Yang, G. W., Wu, H. T., Dong, J., Sun, L. L., Xue, Y.,
Zhao, A. L., Gao, Y. S., Zhu, J. P., Chen, S. X., Yao, Z. J., Wang, Y.,
Lu, W. C., Qiang, B. Q., Wen, Y. M. and Hou, Y. D.
Direct Submission
Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry
of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P. R.
China
FEATURES
source
1. location/Qualifiers
/organism="Shigella flexneri 2a str. 301"
/mol_type="genomic DNA"
/strain="301"
/serotype="2a"
/db_xref="taxon:198214"
148. .1440
/gene="xseA"
/locus_tag="SF2555"
148. .1440
/gene="xseA"
/locus_tag="SF2555"
/note="Residues 1 to 430 of 430 are 97 pct identical to
residues 27 to 456 of a 456 aa protein from Escherichia
coli K12 ref: NP_417004.1"
/codon_start=1
/transl_table=11
/product="exonuclease VII, large subunit"
/protein_id="AA0405.1"
/db_xref="GI:24052929"
/translation="MGQWISGEISNFTOPASGHYFTLKDDTAQVRCAMFRNSNRV
TTPHGGQVIVRANTIVYRGDYOIVESMPAGEGLQKYEQLKAKLOAECLD
QQKRPSPAHGCVGITSTKTAALDILHYKRPPSLPVITPTSYGDDPQPGIV
RALEIANRVECPVALVGRGGSLEIDMSFNDERVARAFASRIPVASAGHETDVTI
ADVAADLRAPTPAAAEVVSRRNQELRQVSRRLERLMAEDYLLNRRRRPQIHHR
LQOQHPLRLAQOTMLERLQKMSFALENQLKACQCOOQLTQLNQQNPQPKIART
OTRIQOLEYRLAIELRAQLSATRERFGNAVTHLEAVSPILTAGVSVTTATGAVLX

gene
complement(1437. .1688)
/gene="yfgJ"
/locus_tag="SF2556"
complement(1437. .1688)
/gene="yfgJ"
/locus_tag="SF2556"
/note="Residues 1 to 83 of 83 are 97 pct identical to
residues 1 to 83 of a 83 aa protein from Escherichia coli
K12 ref: NP_417005.1"
/codon_start=1
/transl_table=11
/product="orf, conserved hypothetical protein"
/protein_id="AA04056.1"
/db_xref="GI:24052930"
/translation="MNVEGMATGIGIMELHCKCOHVLDDNGHARCPGGEFIEKKA
LPDCHQPLQYLKACGAVYFCOHGILSKKRVVLA"
complement(1721. .3232)
/gene="yfgK"
/locus_tag="SF2557"
complement(1721. .3232)
/gene="yfgK"
/locus_tag="SF2557"
/note="Residues 1 to 503 of 503 are 99 pct identical to
residues 1 to 503 of a 503 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_289064.1"
/codon_start=1
/transl_table=11
/product="putative GTP-binding factor"
/protein_id="AA04057.1"
/db_xref="GI:24052931"
/translation="MRCLMITYKNEALMNPVVALVGRPNVSKSTLFNRLTRFDALVA
DPPGLRDRKYEALIEGREFICIDTGGIDGDEPPEVVAQFPAEENGEEDDPQSLPIKL
DARGMLPDEPAIAXHLSREKPTCPIDTGGIDGDEPPEVVAQFPAEENGEEDDPQSLPIKL
GRCVSLIETHLPWMEELAPQEVDEDEVAQFPAEENGEEDDPQSLPIKL
ALVGRPNVSKSTLFNRLIEGRVYVDMGTRDSIYIDMERDGRVLIIDTAGVKR
GRITDAVEKSVITKLQALIDANVVMVIDAREGISDQSLILGFLINSGRSIVIVN
KMDLSQAEQVETPLRGLFIDPARVHTISALHSGGVNLFESVRAVDSSTRV
GTSMLTRITMVAEDHOPPLVGRVYKLYAHAGGVNPPIVINGQVDDLPDSYKRY
LMNYPRKSLDWNQSPRIOPKSGENPYANKRTLTPTQNRKKRLMKHMKK"
complement(3311. .4489)
/gene="yfgL"
/locus_tag="SF2558"
complement(3311. .4489)
/gene="yfgL"
/locus_tag="SF2558"
/note="Residues 1 to 392 of 392 are 99 pct identical to
residues 1 to 392 of a 392 aa protein from Escherichia
coli K12 ref: NP_417007.1"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="AA04058.1"
/db_xref="GI:24052932"
/translation="MQLRKLLPLGLSLVTLISGCSLFNSEEDVVKMSPLPTVENQFPT
TMASTVSQSGIGNFNSLHPALADVVAARAGLVNLAADGKEIYSLSAEDG
WRSKEPALISGCVITSGHYVYGSERKAOYVAANTGTAQNTKVAAGELASPVSDG
LVYHITSNQLQALNBADGAVKRTVNLDPSPSLSGESAPATAFGAAVYGGDNGRISA
VLMEOGQMIMQORISQATSTEDIRISDVDTTPVNVNGVFLATNGNLTALDNRGQ
IMMKELGVDNFIVDGNRIYVDNDNRVMAITIDGGITLMAOSDLHLRLTSPVLYN
GNLVVDSRGVLAHVINVEDGRFVAQOKVDSGFGTEPVAADKLLIOAKDGVYSITR
"
complement(4500. .5120)
/gene="yfgM"
/locus_tag="SF2559"
complement(4500. .5120)
/gene="yfgM"
/locus_tag="SF2559"
/note="Residues 1 to 206 of 206 are 98 pct identical to
residues 1 to 206 of a 206 aa protein from Escherichia
coli K12 ref: NP_417008.1"
/codon_start=1
/transl_table=11

```
/protein_id="AA081323.1"
/db_xref="GI:26109121"
/translation="METPNSLEWVSPLLGLVLFVPMVLAGFIDSIAGGGGLTIPALM
MAGMSRANALATKTLQACGSGISATYIFIRKRVYSIDQELNAMPVSGMSALLIVG
YVADVRQTLPLTIVCTGLFLMPLGEGEDQKRYGFLPALINGCGVGFDEGFEF
PAAGSFYALAFTVLCCFNLAKTAKTAHAKLNAATNSIGLLFLIGKRVATGFMVLVG
QFGLAMGSRVLVSKQKLRPMIVIVSAVMSAKLYDSHGEILHMLGN"
complement(5338..6162)
/gene="mepa"
/locus_tag="c2874"
/locus_tag="c2874"
/complement(5338..6162)
/gene="mepa"
/locus_tag="c2874"
/locus_tag="c2874"
/BC_number="3.4.99.-"
/function="enzyme; Murein sacculus, peptidoglycan"
/note="Escherichia coli K-12 ortholog: D2328; Escherichia
coli O157:H7 ortholog: z3591"
/codon_start=1
/transl_table=1
/product="Penicillin-insensitive murein endopeptidase
precursor"
/protein_id="AA081324.1"
/db_xref="GI:26109122"
/translation="MNKTAIALALASASLATPWOKITOPVPSAQSISGSPNGC
IVGADTLPIQSEHYQVWRTDQRYFGHPLDMFIDRSQVSNLGMGTVLIGMGMPA
GSRFNGSHASHQGLDVDFILOPKRTMSAOLIPALDLSRDSGHVVPALMKPEI
FSI.IKLAODKDVTRIFVNPATKQOCCLDAGTDRDMLRKVRPGRHRAHVRICPA
DSLECEDQPLPPEDGCGAELQSWFEPKPKTTKPEKTPPLPPSCALLDEHVI"
complement(6166..7254)
/gene="aroc"
/locus_tag="c2875"
/locus_tag="c2875"
/complement(6166..7254)
/gene="aroc"
/locus_tag="c2875"
/locus_tag="c2875"
/BC_number="4.2.3.5"
/function="enzyme; Amino acid biosynthesis: Chorismate"
/note="Escherichia coli K-12 ortholog: b2329; Escherichia
coli O157:H7 ortholog: z3592"
/codon_start=1
/transl_table=1
/product="Chorismate synthase"
/protein_id="AA081325.1"
/db_xref="GI:26109123"
/translation="MMAGNTIGOLFRTTFEGESHGLAGCIVDGPVPIPLTEADLOH
DURRRRGTSRYTQRRREPQVKLSVFGSVTGTGILIENTDROSDYSATIDV
FRGHADYTYEQKIGLDYRGGRSSARERVAAGAIKAKTUAERFGLEIKCLTQ
MGDIPLEIKDMSOYEONPFCPPDKIDALDELRLAKSGDSIGAKVTVAAGVPA
LGSEVFPRLADIADIALMSINAVGVEIGDFVALRGSQNRDEITKDFQSNHAG
ILGSSGQOI IAHMALKPTISITVPGRTINRFGEVEVMTKGRHDPVCIGIRAVPIAE
AMAIYVMDHLRQANADYKTDIPRM"
complement(7286..8551)
/gene="yfcB"
/locus_tag="c2876"
/locus_tag="c2876"
/complement(7286..8551)
/gene="yfcB"
/locus_tag="c2876"
/locus_tag="c2876"
/BC_number="2.1.1.72"
/function="putative"
/note="Escherichia coli K-12 ortholog: b2330; Escherichia
coli O157:H7 ortholog: z3593"
/codon_start=1
/transl_table=1
/product="Hypothetical adenine-specific methylase yfcB"
/protein_id="AA081326.1"
/db_xref="GI:26109124"
/translation="MAGIGLFLNQLYMHFADEFTLSVDNRVLLNLASPRHOLAKOR
LVLLAKCCPLFHSISNPFYVPTSTIDRFKIPSPRVNADLSSWCTSRN
YCHALREYVDKIFVDEAVNELIOTIDMLRWSVSPFSAANIWGHGNDPDEAVOLV
LBSLYLIDIDEDMRTARLTSSSEKRIYERIRVNRIRIYAVYLTNANACGHEFYD
BRVAVPSPIGELINNKFAGLISKQPHILDMCGSCGIALACAYAPPEABDVADLS
PDALVAEONTEERGLIHNVPIRSDFRDLPKVOYDLIVNPPYVAEEMSDPNKY
RHEPELIGASGTDLKLTIRILGNADYLDADVGLICEVNSVHLMEOYDPVPTWL
```

```
Alignment Scores:
Pred. No.: 1,47e-132 Length: 300099
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-921-992-50 (1-372) x AE016764 (1-300099)

QY 1 MethiAsnGlnA1AProIleGlnA1AArgLysSerThrArgIleTyrValGlyAsnVal 20
Db 188410 ATGCATTAACCGAGCTCCAAATTCACCTAGAAATCAACCGATTTACCTGGGAATGTG 188351

QY 21 ProIleG1A5pGlyA1AProIleA1AValG1A5SerMetThrAsnThrArgThrAsp 40
Db 188350 CCGATTGGGAGTGTCTCCATCGCGTACACTCATACCAATACCGCTACGACAGAC 188291

QY 41 ValGluA1AThrValA5nGlnIleLysA1A5eGluA1A5ValG1A1A5pIleValArg 60
Db 188290 GTCGAAGCAACGGTCAATCAATCAAGCGCTGGAGCGTGGCTGATATCGTCGT 188231

QY 61 ValSerValProThrMetAspA1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1 80
Db 188230 GTCTCCGTACCGACGATGACCGCGCAGAACGCTTCAAACTCATCAACGACGTTAAC 188171

QY 81 ValProLeuValA1A5pIleHisPheAspTyrArgIleAlaLeuLysValA1A1A1A1 100
Db 188170 GTCCCTCGTGTGCTACATCCATTCGATATCGCATTCGCTGAAAGTACGCAATAC 188111

QY 101 GlyValA5pCysLeuArgIleAsnProG1A5nIleG1A5nGluA1A1A1A1A1A1A1A1 120
Db 188110 GCGGTGATTTGCTGATTTAACTCGGCAATATCGTAAAGCGCTATTCGATG 188051

QY 121 ValValA5pCysA1AArgAspLysAsnIleProIleArgIleGlyValA5nA1A1GlySer 140
Db 188050 GTGGTTGACGTGTGCGCGCGATTAACCAATTCATTCGATTCGTTAATGCGGATCG 187991

QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnA1A1A1A1A1A1A1A1 160
Db 187990 TTGAAATAAGATTCGAGAAAGATGAGGAAACCAACGCGGAGCGTTCCTGGAATCT 187931

QY 161 AlMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
Db 187930 GCATCGCTCATTTGATCATTCGATCGCTGAATTCGATTCGATTCGATTCGATTCGATTC 187871

QY 181 LysAlSerAspValPheLeuA1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1 200
Db 187870 AAAGCTGTGACGCTTCCTCGCTGTGAGTCTTATCGTTGCGCAAAACCAATCGAT 187811

QY 201 GlnProLeuHisLeuGlyIleThrGluA1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1 220
Db 187810 CACCCGCTGATTTGAGGATCACCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 187751

QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
Db 187750 GCTATTGTTTAAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 187691

QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 187690 GCGGCGATCGGATCGAAGAGATCAAAAGTGGTTTCATTTTGAATGCTGCGGATC 187631

QY 261 ArgSerArgGlyIleLeuPheIleA1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1 280
Db 187630 CGTTGCGGCGGATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 187571

QY 281 IleGlyThrValAsnA1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1A1 300
Db 187570 ATCGATACGTTTAAAGCGCTGCGGACCAACGCTGGAAGATATCATCTCGATGGAAGCTT 187511

QY 301 SerIleIleGlyCysValValA5nGlyProGlyGluA1A1A1A1A1A1A1A1A1A1A1A1 320
Db 187510 TCGATTATCGCTGCGTGGAAATGCGCCAGGTAGAGCGCTGATTTCTACATCGCGCTC 187451
```

QY 301 SerIleIleGIcYsvaIValaangIyProGIyGluIalaleuValserThIleuGIyVal 320
Db 279446 TCGATTATCGCGTGGCGTGAATGCGCCGGGTGAGGCGCTGGTTCTACCTCGGGGTC 279387
QY 321 ThGIyGIVaAnIyIyIySserGIyLeuTYrGIyIaAspGIyValIArgIyIaAspATgLeuAsp 340
Db 279386 ACCGGCGGCAACAGAAAGCGGCTCTTATGAAGATGCGCTGGCAAGACCGCTTCGAC 279327
QY 341 AsnaAspMetIleAspGInleuGluIaAlaArgIleArgIalalyIaIaSerGIyIleuAsp 360
Db 279326 AACACGATATGATTCACACGCTGAGGACCGCATTCGTCGGAAGCCAGTCACTGGAC 279267
QY 361 GluIaIaArgArgIleAspValGInIaIyIyIyS 372
Db 279266 GAAGCGCGTGAATGCGTTCAGCAGGTTGAAAAA 279231
RESULT 10
AE016764/c 300099 bp DNA linear BCT 24-DEC-2002
LOCUS Escherichia coli CFT073 section 10 of 18 of the complete genome.
DEFINITION AE016764 AE014075
ACCESSION AE016764.1 GI:26109116
VERSION
KEYWORDS
SOURCE Escherichia coli CFT073
ORGANISM Escherichia coli CFT073
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 300099)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
Extensive Mosaic Structure Revealed by the Complete Genome Sequence
of Uropathogenic Escherichia coli
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
12471157
2 (bases 1 to 300099)
Welch,R.A., Burland,V., Plunkett,G.D. III, Redford,P., Roesch,P.,
Rasko,D.A., Buckles,E.L., Liou,S.-R., Boutin,A., Hackett,J.,
Stroud,D., Mayhew,G.F., Rose,D.J., Zhou,S., Schwartz,D.C.,
Perna,N.T., Mobley,H.L.T., Donnenberg,M.S. and Blattner,F.R.
Direct Subcloning
Submitted (20-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
LOCATION/Qualifiers
1. 300099
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
complement(163..1383)
/gene="fabB"
/locus_tag="C2869"
complement(163..1383)
/gene="fabB"
/locus_tag="C2869"
/RC_number="2.3.1.41"
/function="enzyme; Fatty acid and phosphatidic acid
biosynthesis"
/note="Escherichia coli K-12 ortholog: b2323; Escherichia
coli O157:H7 ortholog: z3586"
/codon_start=1
/transl_table=11
/product="3-oxoacyl-[acyl-carrier-protein] synthase I"
/protein_id="AA081319.1"
/db_xref="GI:26109117"
/translation="MKRAVITGIGIVSIGNNQOEVLASRGSGITPSOELKDSGM
BSHWGNTLDTTGLIDRKVAFMSDASTVAFISMEOAIAADAGISPEAONPRGLI
ASSGGSPPFQFGADAMGPRGLKAVGVYVTKAMASVSKLATPFPIKHVNSIS
SKATSAHCTIGNAVEQIQLKODIVFAGGEEELCEWACEFDMGLSLSTKYNDTPKKA
SRTYADRGDFVIAGGGMVVELEHAGAHYAEIVGYGATSDGADNVAPEGG
ARCKMAMHAGVDTPIDYNSHGTSTPVDVKELAIAREVFGDKSPAIATAMGHS
LGAAGVQEAIVSLMLEHGFIAPSINIEILDQAGLINTVETDRELTVWNSNGCF

gene
1482..3548
/gene="yfcX"
/locus_tag="C2870"
1482..3548
/gene="yfcX"
/locus_tag="C2870"
/function="putative"
/note="Escherichia coli K-12 ortholog: b2324; Escherichia
coli O157:H7 ortholog: z3587"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yfcX"
/protein_id="AA081320.1"
/db_xref="GI:26109118"
/translation="MRSLTIAKIEPPRYRTYVKKHYSIQANLEFNAEGPVSDF
DDVYFSNDGLEETRYVFLGSHLEKRPFEHPHPLFVAVESGEGTGLNLTLMQARDQ
FREARPOLOLRFHISEKPLRADLALAHQWDELPAEAOLOAWPLPQGR
LLIDEGHVTLDLMPDINELTQLDLSLNQKVDLDFGAPAKNDMMTOMLFNMA
RLARGSTLATPTAGFVRGRIOEAGFTQKRGGRKREMLCGVNEQTLPPFCSTPM
FMRGNSQOEAIIIGGITASALLSLALRGQVTLYCADA PALGASGRGALYPL
ISKHPDALNRPFSNAFTFARRFDLLPVKFDHDKGVYTLGMDERSQKIAQMLSHDL
PAELVAEAVANAVEQITGVATNCSGITYPQGLCPAELTRNVLELAQQGQIHYOH
QLQDLSRKDDGWLINPAGDPAHTSHVVLVANGHQSISGFSQTSLSLPHYSAVAGVSHIPT
PELAKIKOVLGYDGLFPQNPANOHHCIGASYHSGSEETAYSDQOQNRRLIDCF
PQVNAKKEVDVSGKEARCGVRCATDHLPMGVNVDYDATTIVEYASIAEPKAEVASAP
VYDILFMPAALGSRGLCSRLCAETILAAQMSSEPIPMDSATLAALNPNLRKLKLG
KAVKXG"
complement(3669..3947)
/gene="yfcL"
/locus_tag="C2871"
complement(3669..3947)
/gene="yfcL"
/locus_tag="C2871"
/note="Escherichia coli K-12 ortholog: b2325; Escherichia
coli O157:H7 ortholog: z3588"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yfcL"
/protein_id="AA081321.1"
/db_xref="GI:26109119"
/translation="MIAEFERSLIALIDGMVDHASDDELFAAGYLRGHLTAIALES
GDDHSAQVHTTVSSOLEKAIGAGLSPRDQLVTDVMEHLFQOASQC"
complement(3981..4529)
/gene="yfcW"
/locus_tag="C2872"
complement(3981..4529)
/gene="yfcW"
/locus_tag="C2872"
/function="putative"
/note="Escherichia coli K-12 ortholog: b2326; Escherichia
coli O157:H7 ortholog: z3589"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yfcW"
/protein_id="AA081322.1"
/db_xref="GI:26109120"
/translation="MNSTHYEQLIIEINSCFADEFNTRLKGDDEPIYLPADAEPY
NRIVPAHGFYSAIHRIISHWCJAGARDDLVDPGWYCPDGDQAQTOSQFEDEVKPO
ALDWLFCVAGYPPFVSCDNLGDFEPDVRVVFQRRVHAQVMDYLANGIPERPARFIKA
LQNYHTPELTLEQFPWPEALN"
complement(4529..5338)
/gene="yfcA"
/locus_tag="C2873"
complement(4529..5338)
/gene="yfcA"
/locus_tag="C2873"
/function="putative structure"
/note="Escherichia coli K-12 ortholog: b2327; Escherichia
coli O157:H7 ortholog: z3590"
/codon_start=1
/transl_table=11
/product="Hypothetical protein yfcA"


```

YLGSVASAAIAVGTVDFGWDGFMVMIGSILAVILLIVMIGEKRRHOLLQKRN
GG"
4420..6048
/gene="ECs3126"
4420..6048
/gene="ECs3126"
/notes="similar to GUPA_ECOLI gi|1788574 percent identity
100 in 542 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="anaerobic sn-glycerol-3-phosphate dehydrogenase
large subunit"
/db_xref="GI:13362596"
/translation="MKTRDSSQSDVILIGGAGAGIARPCARGLRVILVERHDAT
GATGRNGLHSGARVAVTDASRECISSNQLKRIARICVPTNGLFTLEDDIS
FOATFIRACEAGISABAIIDPOQARIIEPVNVALIGAVRPGDVDPFLITAMLD
AKHGAVILTAHEVTGLIREGATVCGVRVNHLTGEQALHAAVVAAGIMQCHIAE
YADLRIMFPKAGSLIMDRINOHVNRKPSDAILVSGDTISLIGTTSRIDYN
EIDNRVTAEVDIILREGKLA PVMAKTRILRAYSQVRLVSDDDPSGRVNSGIV
LIDHARBDGLDRIITIGKLTMRPLMAEWTADPACVKIENPRTADALDGSDEP
AEVTLKRVISLPAPLRSAYVRHGDRTPAWLSGRHLRLVCECAVTAEEVAYEN
LVNLSLIDLRRTRVGTCGELCAGRAAGLLQRFVNTTSAOSIQSLTFINRWKG
VQPIAWGDALRESEFTRWVYQGLCGLEKEKQDAL"
6038..7297
/gene="ECs3127"
6038..7297
/gene="ECs3127"
/notes="similar to GUPB_ECOLI gi|1788575 percent identity
98 in 419 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="anaerobic sn-glycerol-3-phosphate dehydrogenase
membrane anchor subunit"
/protein_id="BAB36550.1"
/db_xref="GI:13362597"
/translation="MRPRTIMGGGAGLIGLOLQKHGRCAIVTRGOSALHPSGGS
LDLSHPDQAVTDIHSGLESARQAPARPYSLGQRVLDAACQALIASGQOL
QSSVELPHQRIITPLGLRSTWSSPEVPVPLPAKCI CVGIGLMDFOHLLAAALR
ELDLAEVETAEIELPELDVLRNNATEFNAVIAFLDNEENWPLLDLALPVANTCMI
LMPACFLADDKLMRWLNELPCSLMLPTLPPSVGLIRLQNLQRFVQGVNMG
DEKVKTCCKGVNMEIWRNHADILPRPREVAIVSGSFSGGLVAEENGIREFIIGD
VLQDARGEWYKGFAPAPQWQOQRYTDTLAPSQAGCTIENLFTIGSVLGFPDPIA
QCGGCVCAVSALHAAQQLKQRAAGQO"
7294..8484
/gene="ECs3128"
7294..8484
/gene="ECs3128"
/notes="similar to GUPC_ECOLI gi|1788576 percent identity
100 in 396 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="anaerobic sn-glycerol-3-phosphate dehydrogenase
K small subunit"
/protein_id="BAB36551.1"
/db_xref="GI:13362598"
/translation="MNDTSFENCIKCTVCTTACPVSRVNFPGYKQAGPGRRLK
DGALEYALKYCNCKCEVAECPSDYKIGIQRARAKYPTPSLRNPFVLSHIDMG
SVTFPAPIVNTATSLKPVROLDAALIKHRLPLPSYSGTRRWVVAADQAYK
DOVAFFHGFVNYNHQLGKDLIKVLANMGTGLSKKCCGVPIANSPADAKO
AIRNVEISIRAVGKGIPIVATISSTCTFALRDEYPRVIANDKGLRDLIELATRMIR
KLDBGKTLPLKPLKAVYHPCHEKMGWLTLELRNI PGLLETVLDSQCCGLAG
TYGFKENYPTQAIGAPLRFQIRESGADVVDCECKQIEMSTSLRCEHPITLLA
QALA"
8477..9603
/gene="ECs3129"
8477..9603
/gene="ECs3129"
/notes="similar to YPAD_ECOLI gi|1788577 percent identity
94 in 303 aa but differs at C-ter (Conserved in E.coli

```

```

Alignment Scores:
Pred. No.: 1,45e-132 Length: 296827
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0

US-09-921-992-50 (1-372) x AP002561 (1-296827)
QY 1 MethiAsnGlnAlaProIleGlnArgLysSerThrArgIleTyrValGlyAsnVal 20
Db 280346 ATGCATACACGAGCTCAATTCACAGTAAGAAATCAACAGTATTACCTGGGAATGTC 280287
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 280286 CCGATTGGCGAGTGTCTCCATCGCGCATGATCATCAACCAATACGCGTACGACAGAT 280227
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
Db 280226 GTCGAAGCAACGCTCAATCAATCAAGGCGCTGGAACGCGTGGCGCTGATATCGTCCG 280167
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
Db 280166 GTCTCCGTCACCAAGATGACGCGGAGAGAGCTTCAAACTCATCAACAGCAGTTAAC 280107
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValAlaGluTyr 100
Db 280106 GTCCCGCTGATGGCTGACATCCACTTCGATATCGGCTGGAAGTACGGAATAC 280047
QY 101 GlyValAspCysLeuArgIleLeuAsnProGlyAsnIleGlyAsnGluLysIleArgMet 120
Db 280046 GGGGTGATTGCTGTGGTATTAACTCGCAATGCTGAATGAAAGCGTATTCCCATG 279987
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAlaAsnIleGlySer 140
Db 279986 GTGGTTGACTGTGCGCGCATTAATAAATTCGATCCGATTGCGTTAACCCGGATCG 279927
QY 141 LeuGluLysAspLeuGlnIleLysTyrGlyGluProThrProGlnAlaLeuLeuGlnSer 160
Db 279926 CTGGAATAAAGATCTGCAAGAAAGTATGGCAACGACCGCGGCGTTCGTGAATCC 279867
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnAspGlnPheLysValSerVal 180
Db 279866 GCCATCGCGCATGTTGATCATCTCGATCCGCTGAACCTTGATCATCAAGTCAAGCAGCTG 279807
QY 181 LysAlaSerAspValPheLeuAlaValGlnSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 279806 AAAAGCTCTGACCTCTCTCGCTGTGAGCTTATATGCTTGTGCAAAACAAATCAT 279747
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValArgSerGlyValAlaLysSer 220
Db 279746 CAGCGCTGATCTGGGATCAACGAAAGCGGGTGGCGCGGACGAGGAGTAAATCC 279687
QY 221 AlaIleGlyLysGlyLeuLeuLeuSerGlyIleGlyAspThrLeuArgValSerLeu 240
Db 279686 GCCATTGCTTTAGTCTGCTGCTGCTGCTGAAGGATCGGAGACACGCTGCGCATACCG 279627
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 279626 GCGGCGGATCCGCTGCAAGATCAAGATCGGTTTCGATATTTTGAATCCGCTTCGATC 279567
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnLysPheAspVal 280
Db 279566 CGTTGCGAGGATCAACTTCAATCGCTGCGGACCTGTTCCGTCAGGAATTTATGTT 279507
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
Db 279506 ATCGGTACGATTAAACCGCTGAGCAACCGCTGGAAGATATCATCATCTCCGATGACGTT 279447

```


Db	364	ATCGTAGCGTTAAACCGCTGGAGCAACCGCTGGAAAGATATCATCTCCATGGACGTT	305
Oy	301	SerTIIeIEgIcYsValValAsnGlyProGIgIuAlaLeuValSerThIreuGIyAl	320
Db	304	TGCATTATTCGCTGCTGGTGTGTAATGCCACGAGTGAGCGCTGGTTCTACACTCGCGCTC	245
Oy	321	ThrGIgIyAsnIlyAsnIlySeSeGIyLeuIYrGIyAspGIyValArgIlyAspArgIleuAsp	340
Db	244	ACGGCGGCGCAACAAAGAAACCGCGCTCTATGTAAGATGGCGTGGCAAGAACCGCTGTGAC	185
Oy	341	AsnAsnAspMetIleAspGlnIleuGlnIuAlaArgIleArgAlaIyAlAsenGlnIleuAsp	360
Db	184	AACAACGATATGATGATGACCACTGAGTGAAGACCATTCGTGGCAAGAACCATGACTGGAC	125
Oy	361	GIuAlaArgArgIleAspValGInGInValGIuIyS	372
Db	124	GAGCGCGCTCGAATTGACGTTGACAGGTTGAAAA	89
RESULT 9	AP002561/c		
LOCUS	AP002561	296827 bp	DNA linear BCT 24-APR-2003
DEFINITION	Escherichia coli O157:H7 DNA, complete genome, section 12/20.		
ACCESSION	AP002561	BA000007	
VERSION	AP002561.1	GI:13362592	
KEYWORDS	Escherichia coli O157:H7		
SOURCE	Escherichia coli O157:H7		
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.		
REFERENCE	1	Makino, K., Yokoyama, K., Kubota, Y., Yutsudo, C.H., Kimura, S., Kurokawa, K., Ohishi, K., Hattori, M., Tatsuno, I., Abe, H., Iida, T., Yamamoto, K., Ohnishi, M., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C., and Shinagawa, H.	
AUTHORS	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 gene of the enterohemorrhagic Escherichia coli O157:H7 derived from the Sakai outbreak		
TITLE	Genes Genet. Syst. 74 (5), 227-239 (1999)		
JOURNAL	2	Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.	
MEDLINE	20198780		
PUBMED	10734605		
REFERENCE	3	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an Escherichia coli K-12 strain MG1655	
AUTHORS	20557356		
JOURNAL	11108008		
MEDLINE	20564182		
PUBMED	1111050		
REFERENCE	4	Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C. and Shinagawa, H.	
AUTHORS	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli O157:H7 strain derived from the Sakai outbreak		
TITLE	Gene 258 (1-2), 127-139 (2000)		
JOURNAL	20564182		
MEDLINE	1111050		
PUBMED	1111050		
REFERENCE	5	Yokoyama, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Hayashi, T., Tobe, T., Iida, T., Takami, H., Honda, T., Sasaki, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.	
AUTHORS	Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12		
TITLE	DNA Res. 8 (1), 11-22 (2001)		
JOURNAL	21156231		
MEDLINE	1158796		
PUBMED	1158796		

REFERENCE	5 (bases 1 to 296827)
AUTHORS	Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.
TITLE	Direct Submission
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/ , Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)
COMMENT	genome project.
FEATURES	Location/Qualifiers
source	1..296827
	/organism="Escherichia coli O157:H7"
	/mol_type="genomic DNA"
	/strain="O157:H7"
	/sub_strain="RIMD 0509952"
	/db_xref="taxon:83334"
gene	414..1484
	/gene="ECs3123"
CDS	414..1484
	/gene="ECs3123"
	/note="unknown"
	/codon_start=1
	/evidence=not experimental
	/transl_table=11
	/product="hypothetical protein"
	/protein_id="BAB36546.1"
	/db_xref="GI:13362593"
	/translation="MPNTSIHLSRCNLIQNNKLTQPTETIYKESQOTAKLEPFCDPEFLK SOSQVSTSDSIANLAPPTKASDAIDSKQKIVHSYGFPIRHFVHVITYKK EBNADYEPYLTITPTRGNNBGLSVESKEGTSISRTINANFEANININHVIKEN HNATQPTFKYTSKINSDDPKYKTSKESIKSOLYSQPNQIQTYILIRKNTWKVYK SSSERYDKFNSSKEIKNTYKQIQAQKQNMKYINKLSELIDIDKLAPLINDSE FSYRDYFAYSKQKKNITFTFNSKFFVDPKEKEXGLMVHNSGLVITGKYCEGQTA TOEYDLIRINIFTFPEKLE"
	complement(11708..2784)
gene	/gene="ECs3124"
	complement(11708..2784)
CDS	/gene="ECs3124"
	/note="similar to GLPQ_ECOLI g11788572 percent identity 99 in 358 aa (Conserved in E.coli K-12)"
	/codon_start=1
	/evidence=not experimental
	/transl_table=11
	/product="glycerophosphodiester phosphodiesterase"
	/protein_id="BAB36547.1"
	/db_xref="GI:13362594"
	/translation="MKLTIKLSMAIMSTIWMSSAMADSNKIVIAHAGSGYLE EHTLPKAMAVYAGADLTLEDLVTKDHLVLDHYLDRTVDADRPDRKQDGRG YALDITDEIKSLKFTGTFDENGKQVTPGRPMKSDPRHTFEELEIFVGLNLS STAGNIGIPEIKAFPHHQBQIDAKITLVEIKKYVYGKDDKYVLCQFPADELKR KTELEPMQMDNLVOLIAITDMNETOOKOPDGSVMYNDMPKPGAMKQVAEADQ IGPEPHYLIEBTSQPNIKILNGWYQDAQOKLVVHPYVNSDKLPETTYDNOYLDAI YNKAQVNGLTDPDPKAVFLNKE"
	complement(2789..4147)
gene	/gene="ECs3125"
	complement(2789..4147)
CDS	/gene="ECs3125"
	/note="similar to GLPT_ECOLI g11788573 percent identity 99 in 452 aa (Conserved in E.coli K-12)"
	/codon_start=1
	/evidence=not experimental
	/transl_table=11
	/product="sn-glycerol-3-phosphate permease"
	/protein_id="BAB36548.1"
	/db_xref="GI:13362595"
	/translation="MUSTFKPAPRAKALPPAAEIDPTVYRLRLWOQIFGI PCGYAAYLVVLRKFAFLMPYLIVEGSGFGRDLGFLSISLAVGSKRIMSVSDRSNPRVFLPAGLILIR AAVALFMGFEVPAWTSIAVMPVLLFCGWFQGGWPCGRTWMMHMSOKERGIVSGV MCAAHNVGGGIPLPLFLIGMAWMDHDAALYMPAFCAIYALFAFAMWRDPGSCGLE PIREYKNDYRDDYNEKAQDELTKQIPQCVVLDPKILMIAYIANVPYLRLYGIILMS PTLYKEVHFLALDKSSWATFYEYAGIDGILLGMSMDKVPFRGRGATGVFPMVLVTI ATTIVWNNPAGNPTVDMI CMIVIGFLYGGVMIIGHALBELAPKKAAGTAAAGFTGLGFG

```
CDS      complement (2529..3683)
          /gene="yfgB"
          /function="orf; Unknown"
          /note="f384; 100 pct identical to YFG_ECOLI SW: P36979"
          /codon_start=1
          /transl_table=11
          /product="orf, hypothetical protein"
          /protein_id="AAC75570.1"
          /db_xref="GI:1788865"
          /translation="MSRQVLPENVTTKKCKINLIDLNQKMEFRKDLCEKPRPDQ
          VKMWHYCCDNFDEMTDINKVLGKLEVAELRAPVEEKS SDGTIKWALAVDQ
          RVEYVI PEDRATLCVSSQVGCALCEKFCSTQGFNRLRVSEIIGVWRAKIVG
          AAVTGORPTITVMMGMGEPLLNINVPAMEIMDDFEGSLKRVLTISTGVIPA
          LIDGMIDVALAISLHPNDEIRDEIVPINKYNIETFLAARVYLEKSNAGKRV
          LEYVMDHVNDGTEHAHQALAEIKDTPCKINLIPNPPCAPYGRSNSRIDRFSVL
          MSYFTTVKRTKGGDDIDACGQAGADVIDRTKTRTKRKGSAITDIKAV"

promoter  complement (3778..3806)
          /note="factor Sigma70; predicted +1 start at 2642391"
          /codon_start=1
          /transl_table=11
          /product="nucleoside diphosphate kinase"
          /protein_id="AAC75571.1"
          /db_xref="GI:1788866"
          /translation="MAIERFSLIKRNAVAKVIGNIPARFEAGKTYCTKMLTLY
          EQARGFAEHDGKPFPGDVEFMTSGPIVSVLEGNAVGRHDDLGAITNPANALAGT
          LRADVADSLTENGTHGSDSVESAAREIAVFFGEGEVCPRTR"

promoter  complement (4322..4350)
          /note="factor Sigma70; predicted +1 start at 2642935"
          /codon_start=1
          /transl_table=11
          /product="putative peptidoglycan enzyme"
          /protein_id="AAC75572.1"
          /db_xref="GI:1788867"
          /translation="MPRLITKRCWITLAAAPLFLAAMGADKMLPLLHEVNPARY
          VVAQDCTPLMRPADAGIWRPVITIEDVSRPYLEALINVEDRFMGHPGNPESVARA
          AMODLTSGRVISGSGITLMQVARIIDPHPTFGKIRQLRMLALFEMHLSKRILLIY
          LNAEPGGTLOGIGASMAWYLGKSPANLSSSEAMLAIVLPQASRLRPBMRPAA
          RNKVLREMAVQCVMSRBOYSESEPEITLAPROMPOLAPLFSRMLGKSKSDITTL
          DACLQRLELAQNKGRLEPPRSSLMTIVDHTDKVRGVGSDIANDSRFSHVMSEA
          NVSLNPVAVOVLAEVGRKFAKRLNVGILPLYPNGAIPNLILIGGAGAKLEDMA
          AYAPAFHKGAKLRLQDDPLLERPLMSGAMIIRIIMADQAPLPSALRVAVPL
          AKMTGSGYGRMAIGVNAVYIGIWTGPGDGTGVGOGFASAVPLINOVNITLS
          RSNLNPEDRPNSVTGVCVCPGGSGILPEDGCRRLNMLIDGQPTLILPEEG
          INCIIRPIMLDENGRKVAADCPAROBMTWPLPLEPMPAPERRAVRIPASTSCP
          PEGHDAQPLQLTGVBDGALIKRLPGAARETTLPLQSSGGAGGERWFLNEPLTERGRN
          VTLLHTDKGVDYQLNMDVGOALIVKFKVMQ"

gene      complement (6726..11687)
          /gene="b2520"
          /complement (6726..11687)
          /function="orf; Unknown"
          /note="f1653; f1653"
          /codon_start=1

CDS      complement (6726..11687)
          /gene="b2520"
          /complement (6726..11687)
          /function="orf; Unknown"
          /note="f1653; f1653"
          /codon_start=1

CDS      complement (6726..11687)
          /gene="b2520"
          /complement (6726..11687)
          /function="orf; Unknown"
          /note="f1653; f1653"
          /codon_start=1
```

```
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC75573.1"
/db_xref="GI:1788868"
/translation="MKRLRVAAQMLMALAGCDNNDNAPPAVKQAPSEVTYKAASSEN

Alignment Scores:
Pred. No.:      2,366-134      Length:      13176
Score:          1666.00      Matches:      372
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:      100.00%      Indels:      0
DB:              Gaps:      0

US-09-921-992-50 (1-372) x AB000338 (1-13176)

QY      1 MethIASnglnAlaProIleGlnArgLYsSerThrArgIleTyrValGlyAsnVal 20
DB      1204 ATGCATTAACCGAGCTCCAAATTCACGTAGAAATCAACAGTATTACGTTGGGAATGTC 1145

QY      21 ProIleGlyASpGlyValaProIleAlaValGlnSerMetThrAsnThrArgThrTrpAsp 40
DB      1144 CGATTGGCGATGGTGGCTCCATCGCGGACATCATACCAATACGGGTAGACAGAC 1085

QY      41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyValaAspIleValArg 60
DB      1084 GTCGAAGCAACGGTCAATCAATCAAGCGCGTGAACGGCTTGCGCTATATCGTCCT 1025

QY      61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValaAsn 80
DB      1024 GTATCCGTACCAAGCATGACGACCGGCAAGAGCTTCAAACTCATCAACAGCAGGTTAAC 965

QY      81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValaAlaGluTyr 100
DB      964 GGCGCGCTGTGGCTGACATCCATCGATCGATTCGATTCGATTCGATTCGATTCGATTCG 905

QY      101 GlyValaAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGlnIleArgIleArgMet 120
DB      904 GGGCTGATTCGTCTGGTATTAACTTGCATATCGTGAATGAGAGCGTATTCGATTC 845

QY      121 ValValaAspCysAlaArgAspLysAsnIleProIleArgIleGlyValaAsnAlaGlySer 140
DB      844 GTGGTTGACTGTGGCGGCGATTAACATTCGATTCGATTCGATTCGATTCGATTCGATTCG 785

QY      141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
DB      784 CTGGAATAAAGATCTGCAGAAAAAGTAGCGGAACCGACCGCGGCTTGCATCT 725

QY      161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAsnPheAspGlnPheLysValSerVal 180
DB      724 GCCATCGCATGTTGATCATCTCATCTCGATCGCTGACCTTGATTCAGTCAAAAGTCAGCGTG 665

QY      181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB      664 AAAGCGTCTGACGTCCTCTCGCTGTTAGTCTTATCGTTCTGCAAAACAGATCAT 605

QY      201 GlnProLeuHisLeuGlyLerThrGluAlaGlyValaArgSerGlyValaLysSer 220
DB      604 CAGCGCTTCATCTGGGATCAACGAGCGGATGCGGACGCGGACGCGGACGCGGACGCGGATC 545

QY      221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
DB      544 GCCATTGGTTAGTCTGCTGCTGCTGCTGAAGCATCGGCAACGCTGCGGATTCGCTG 485

QY      241 AlaAlaAspProValGluGluIleLysValaGlyPheAspIleLeuLysSerLeuArgIle 260
DB      484 GCGGCGGATCGCGGATCAAGATCAAGACGCGGATTCGATTTGAAATCCGCTGCGATC 425

QY      261 ArgSerArgGlyTyrLeuAsnPheIleAlaCysProThrCysSerArgGlnIlePheAspVal 280
DB      424 CGTTGCGGAGGATCAACTTATCGCTGCGGACCTGTTCCGTCAGGAATTTGATGTT 365

QY      281 IleGlyThrValAsnAlaLeuGlnIleArgLeuGluAspIleIleThrProMetAspVal 300
```

Db 424 CGTTCGGAGATACACTTCATCGCTCTCCGACCTGTTCGGCTCAGCAATTTGATGTT 365
 Qy 281 ILGGLThrValAsnAlaLeuGlnArgLeuGlnAspIleIleThrPrometAspVal 300
 Db 364 ATGGATACATTAACGGCTGAGCAACGGCTGAAATATCATCTCCGATGAGAGTT 305
 Qy 301 SerIleIleGlyCysValValAsnGlyProGlyGlnAlaLeuValSerThrLeuGlyVal 320
 Db 304 TCATATTATCGCTCGCGTGGTAATGGCCCGGGTGAAGCGCTGGTTTCTACACTCGGCTC 245
 Qy 321 ThG[V]GValanlyblybSerGlyLeuTyTGluAspGlyValArglyAspAlaGluAsp 340
 Db 244 ACCGGCGGCAACAGAAACCGGCTCTTAAGAAGATGCGCTGGCAAAAGCCGCTTGAC 185
 Qy 341 AsnAsnAspMetIleAspGlnLeuGluValArgIleArgAlaIalysAlaSerGlnLeuAsp 360
 Db 184 AACACAGATATGATCGACCGAGCTGAGGACCGCATTTGTCTCGAANAAGCATGCTGAGAC 125
 Qy 361 GUAAlaArgArgIleAspValGlnGlnValGlnLys 372
 Db 124 GAAGCGGCTGCAATTCAGCTTCAGACAGTTGAAAAA 89
 RESULT 8
 AEO00338/c 13176 bp DNA linear BCT 01-DEC-2000
 LOCUS Escherichia coli K12 MG1655 section 228 of 400 of the complete
 DEFINITION Genome.
 ACCESSION AEO00338 U00096
 VERSION AEO00338.1 GI:1788862
 KEYWORDS
 SOURCE Escherichia coli K12
 ORGANISM Escherichia coli K12
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE
 AUTHORS 1 (bases 1 to 13176)
 Blatner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
 Riley,M., Collado-Vides,D., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
 Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
 Mau,B. and Shao,Y.
 The complete genome sequence of Escherichia coli K-12
 Science 277 (5331), 1453-1474 (1997)
 TITLE
 JOURNAL 9278503
 MEDLINE 9742617
 PUBMED
 REFERENCE
 AUTHORS 2 (bases 1 to 13176)
 Blatner,F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 3 (bases 1 to 13176)
 Blatner,F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459
 4 (bases 1 to 13176)
 Plunkett,G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
 COMMENT
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
 30332 (e-mail: mark@ambler.gatech.edu). Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by

Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products, all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES

source
 location/Qualifiers
 1..13176
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 /complement(49..81)
 /note="factor Sigma70; promoter hns; documented +1 at
 263862"
 complement(86..1204)
 /gene="gcpB"
 /note="synonym: b2515"
 complement(86..1204)
 /gene="gcpE"
 /function="orf; unknown"
 /note="f372: 100 pct identical to GCPE_ECOLI SW: P27433"
 /codon_start=1
 /transl_table=11
 /product="orf, hypothetical protein"
 /protein_id="AAC75568.1"
 /db_xref="GI:1788863"
 /translation="MHNOAPIORRSTRIVGNNPVGDCAPAIKAVOSMTNRTTDEAT
 VNQIKALERVGADIVRSVPTMDAAEFILIQVNVPIVADIDHPYRIALAKAEVGV
 DCIRINPGRINGBERTIRNVDCARDNITIRIGVNAAGSEKDIQEGEPTPALLES
 ARKHVDHLRLNFDQPKVSKASDVFLAVESYRLAKQIDQPHLQITEGPGASGV
 KSAIGLGLSEIGIDTLRLSLADPVEIKYGFILKSLIRSRGINFACPTCSRO
 EFDVIGTVNALGQRLEDITTPMDVSIIGCVNPGPALSTLGTGNGKSKGLYEDGV
 RKDRLDNNMDILEARIRAKASQDEARIDVQVEK"
 complement(1231..2244)
 /gene="yfgA"
 /note="synonym: b2516"
 complement(1231..2244)
 /gene="yfgA"
 /function="putative membrane; Not classified"
 /note="f337; 100 pct identical to YFGA_ECOLI SW: P27434"
 /codon_start=1
 /transl_table=11
 /product="putative membrane protein"
 /protein_id="AAC75569.1"
 /db_xref="GI:1788864"
 /translation="MNTETKTDHONENLTTCARLRNARBQSGQAVARLCLKVSTV
 RLIEDKPADLASTFTLRGYSYARLVHPEELPGLEKQAPLEAAVAPQSFSL
 GRRKRKRDGMLMTFWLVFVIVIGSGAMWDRKQOEITMAQSSAELESNSBO
 GGSVPINTSTTDPATSTTPASVDVATNTQTPATVAPAVDPQNAVSPSQNV
 DPAATPAATATTPDGAAPLPIDQAGVTPVADPNVLAWNFTADCELEVDATGKLF
 SEMQRKDGMLNTGQAPYTKTGAPAAVIOYQGRKVDLSRFRITNQVARLTLMASOS
 PAQ"
 complement(1288..1316)
 /gene="yfgA"
 /note="factor Sigma70; predicted +1 start at 2639901"
 complement(1360..1389)
 /gene="yfgA"
 /note="factor Sigma70; predicted +1 start at 2639973"
 complement(2529..3683)
 /gene="yfgB"
 /note="synonym: b2517"

gene
 promoter
 promoter
 gene

misc_feature
LRADVADSLTNGCTHGSVSASAREIAVFPGEVCPRTT"
4397..9255
/note="O-island #106: Region of the EDU933 chromosome not homologous to E. coli K-12 MG1655"
gene
complement(4405..4983)
/gene="23782"
complement(4405..4983)
/gene="23782"
/function="orf: Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=1
/protein_id="AAG57629.1"
/db_xref="GI:12516913"
/translation="MLMKEQGTATAIDISQAKWERALLNECLROYGEGWV
SPRTSEVNSRRRLPHYQADVQMTQIQPRLRLHLRIPISAPRPMDQKQCLGA
CMFACEQVPSLENGHQLINDALCNGCNCIACFQAMVELTLLPAKIVLHNRK
VCTQCKSFLTFQONQNCIYQHRXGMRTP"
complement(5259..6080)
/gene="23783"
complement(5259..6080)
/gene="23783"
/function="putative enzyme; Not classified"
/note="Residues 2 to 272 of 273 are 25.26 pct identical to residues 6 to 276 of 286 from GenPept 118 : gi15002126|gb|AAD37317.1|AF13170.8 (Afi35170) dimethyl sulfoxide reductase subunit C [Yersinia pestis]"
/codon_start=1
/transl_table=1
/product="putative dimethyl sulfoxide reductase subunit C"
/protein_id="AAG57630.1"
/db_xref="GI:12516914"
/translation="MHELRPMFTLLQASVGMWLMARHRIITSSRPMRPL
LAFPLAGVGLSTHLGTPLRNMLHLSFSSMSREITFALYALGLSTLVA
LRKGGQQLMLAAVGIIVDFCMAQVYHTSIITQWVTLMLFGVIGLSALVA
VRFISQILPQIDALRNGCVLIALVLRLLVPLWIGDITANMOIATLPHAPLML
GQLKPLITLSWGSVIGMFFAVGGCKNIPALFGSVMLVSEVMLEFVFSIG"
complement(6073..6510)
/gene="23784"
complement(6073..6510)
/gene="23784"
/function="putative enzyme; Not classified"
/note="Residues 4 to 134 of 145 are 61.83 pct identical to residues 57 to 186 of 205 from GenPept 118 : gi11787122|gb|AAC73981.1|AB000191|anaerobic dimethyl sulfoxide reductase subunit B [Escherichia coli]"
/codon_start=1
/transl_table=1
/product="putative dimethyl sulfoxide reductase subunit B"
/protein_id="AAG57631.1"
/db_xref="GI:12516915"
/translation="MHNHFAVYTLVSCHNCADPCTKNCPTMAMHRRPDDGIVRVNT
DKVCGGCGYCAMSCPYGAPQNNQETGMKSKDCIDIQAKQEQICVATCGLGAIKFGP
IDLRKRYGVSIVKGLPDSITQPNLVIKPHQGAKEKSHHA"
complement(6699..9080)
/gene="23785"
complement(6699..9080)
/gene="23785"
/function="putative enzyme; Not classified"
/note="Residues 25 to 792 of 793 are 43.59 pct identical to residues 3 to 784 of 785 from GenPept 118 : gi11787121|gb|AAC73980.1|AB000191|anaerobic dimethyl sulfoxide reductase subunit A [Escherichia coli]"
/codon_start=1
/transl_table=1
/product="putative dimethyl sulfoxide reductase subunit A"
/protein_id="AAG57632.1"
/db_xref="GI:12516916"
/translation="MDTPDEKERITTPQVSRBSFLQATSLITLPIRISSTAKQSDA
SEPTVAPADKVPCTCFDGGKCDIRAMRGVVTQITLDDNLEDPMPIMRACV
RGGRYKRVYHPRLKYPKMRVGKRGKGRISWDEATLLIDNENKRIITQYQPSR
YVAVGTAVXGTSFGDAMARLLNLGTGYLEYHVSIGNTAATATYTYGVAAGSM
DTLDTKLVILMGHNPTETIFGHTNYFYQMKONGTRFIVVDRVSDTVSLADQMP

Alignment Scores:
Pred. No.: 1,986-134 Length: 11521
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-09-921-992-50 (1-372) x AE005481 (1-11521)
QY 1 MethiAsnGluAlaProIleGlnArgValSerThrArgIleTyrValGlyAsnVal 20
DB 1204 ATGCATACCAAGGCTCCAAATTCACAGTAAGAAATCAACGTAATTAAGTGGGAATG 1145
QY 21 ProIleGlyAspGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB 1144 CGGATTGGGAGTGGTCTCCCATCGCGCTGACATGATGACCAATACGCGTACGACAGAT 1085
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyAlaAspIleValArg 60
DB 1084 GTCAAGCAACAGGTCATCAATCAAGGCGCTGGAAGCGGCTGGCCCTGATATCGCCGT 1025
QY 61 ValSerValProThrMetAspAlaIleGluAlaGlyValLeuGlnGlnValAsn 80
DB 1024 GTCTCCGTCGACGATGAGCGGAGAAAGCCTTAACCTCAACAAACGCGATTAC 965
QY 81 ValProLeuValAlaAspIleHisIlePheAspTyrArgIleAlaLeuValAlaGluTyr 100
DB 964 GCGCCGCTGGTGGCTGACATCCACTTCGACTATCGCATTCGCGTGAAGTACGGAATAC 905
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluLysIleArgMet 120
DB 904 GCGGTGATTTGCTGCTGATTTAACTTCGCAATATGTAAGAGCGTATTCGATG 845
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnIleArgSer 140
DB 844 GGGTGACTGCGCGGCGCAATAAACATTCGATCGATTGGCCGTAAACGCGGATCG 785
QY 141 LeuGluValAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGluSer 160
DB 784 CTGGAATAAAGATCTGCAABAAAGTATGGCGAAGCAGCCCGAGCGCTTCGGAATCC 725
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
DB 724 GCATCGCGCATGTTGATCATCTCGATCCCTGMACTTGATGATCAAGTCAAGTCAAGT 665
QY 181 LysAlaSerAspValPheLeuAlaValGlnSerTyrArgLeuLeuAlaLysGlnIleAsp 200
DB 664 AAGGCTCTGACGCTCTCTCGTGTGAGCTTATGCTTGCAAAACAAATCGAT 605
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyAlaValLysSer 220
DB 604 CAGCGCTGATCTGGGATCAACGAGCGGCGCGCGGCGGAGCGGAGGAGTAAATCC 545
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
DB 544 GCATTTGTTTGGTCTGCTGCTGCTGATGAAAGGATGCGGACACGCTGCGGATCATCG 485
QY 241 AlaAlaAspProValGluGluLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB 484 GCGGCGGATCGGTCGAAAGATCAAGTGGTTTGATATTTGAAATCGCTTCGATAC 425
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnLysPheAspVal 280

QY 261 ArgSerArgIylleAenPheilleaAcyProthrCysSerArgIngluPheAspVal 280
DB 781 CGTTCGGAGGATCACTTCATCGCTGCCACCGCTTCGCTCGAGAAATTGATGTT 840
QY 281 ILGLYThrValAenAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
DB 841 ATCGGTAACGGTTAACCGCTGAGCAACGCTGGAGATATCATCACTCCGATGAGCTT 900
QY 301 SerIleIleGlyCysValValAenGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
DB 901 TCATTATTCGCTGCGCTGTAATGCGCCAGGTGAGCGCTGCTTCTACACTCGGCTC 960
QY 321 ThrGlyValAsnIlySerGlyLeuTyGlyAspGlyValAlaArgLysAspArgLeuAsp 340
DB 961 ACCGCGCGCAACAAGAAACCGGCTTATGAAAGATGCGCTGCGCAAGACCGCTTGAC 1020
QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
DB 1021 AACACGATATGATCGACCGACGAGCAACGCGATTCGTGCGAAAGCCAGCTCGAC 1080
QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
DB 1081 GAAGCGGTGAATGACGTTACGACGTTGAAAA 1116
RESULT 6
ECGCE 1697 bp DNA linear BCT 08-DEC-1993
LOCUS E.coli gcpe gene. 1697 bp DNA linear BCT 08-DEC-1993
DEFINITION X64451 S43432
ACCESSION X64451.1 GI:41540
VERSION 6
KEYWORDS gcpe gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 1697)
AUTHORS Parker, J.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-1992) J. Parker, Southern Illinois Univ, Dept of
Microbiology, S I U, Carbondale IL 62901, USA
REFERENCE 2 (bases 1 to 1697)
AUTHORS Baker, J., Franklin, D. B. and Parker, J.
TITLE Sequence and characterization of the gcpe gene of Escherichia coli
JOURNAL FEMS Microbiol. Lett. 94, 175-180 (1992).
COMMENT See also J01629 & M11843.
FEATURES
source
location/Qualifiers
1..1697
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:562"
/sub_clone="PSE411"
/cell_line="JF298"
/clone_lib="lambda transducing phage dguab-2"
/note="ORF"
/codon_start=2
/transl_table=11
/protein_id="CAA45782.1"
/db_xref="GI:41541"
/db_xref="SWISS-PROT:P27434"
/translation="TTTDPATSTPPASVDTTATNTOTPAVTPAPAVDPQONAVSP
SOANVTATPATPATTPDGAAPLPTDAGCTTPVADPAVMTADCMLEVTAT
KGLFSGMQRKDNLMVLTGAPYKLTGAPAAVIOYQGPVLSRFIRTNQVRLTL
NAQSPAQ"
535..1653
/gene="gcpe"
535..1653
/gene="gcpe"
/codon_start=1
/transl_table=11
/protein_id="CAA45783.1"

/db_xref="GI:41542"
/db_xref="SWISS-PROT:P27433"
/translation="MHNQAPLQRRKSTRVYGNVPIGDPAPLAVOSMTTRTQVAT
VNOIKALERVGADYIRVSPYPTDAAEAFELIKQNVNVLADIHPIYRALKVAEYGV
DCLRINPQINIGNEERIKNVVDCAKDNIPIRGVAVSGLEKLDQERYGEPYQALLES
ANRVDHLDRLNEFDQKVSVDVLAESVRLAKQIDQPLHLGKITAGARSGAV
KSAIGGLLSEIGIDGFLRLSLADPVEEIKGFDLKLRSRINFIACPTCSRO
EEDVIGTVALORLEDDITTPMDVSIIGVNVGPGEALVSTIGVGNKSGLYEDGV
RDRPLDNDMDQLSARIRAKSOLDEARIDVQVEK"
BASE COUNT 417 a 446 c 469 g 365 t
ORIGIN
Alignment Scores:
Pred. No.: 1,58e-135 Length: 1697
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-09-921-992-50 (1-372) x ECGCE (1-1697)
QY 1 Meth:AsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyValGlyAenVal 20
DB 535 ATGCATTAACCAAGCTCCAAATTCACGTAAGAAATCAACGATATTACGTTGGAAATG 594
QY 21 ProlIleGlyAspGlyValAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
DB 595 CCGATTGGCGATGGTGTCTCCATCGCGTACAGTCCATACCAATACGGGTAGACAGAC 654
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyValAspIleValArg 60
DB 655 GTCGAAGCAACGGTCAATCAATCAAGCGCGTGGAAAGCGCTTGCCCTATATCGTCCGT 714
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAen 80
DB 715 GTATCCGTACCGACGATGAGCGCGGAGAAAGCTTCAAACTCATCAACAGCAGTTAAC 774
QY 81 ValProLeuValAlaAspIleHisPheAspTyArgIleAlaLeuLysValAlaGluTyR 100
DB 775 GTGCCCTGTGGTGGTACATTCACCTTCGATTCGATTCGCTGAAATAGCGGAATAC 834
QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
DB 835 GCGGTGATGTCGTGCTGTTAACTCGCAATATCGTAATGAAAGCGTATTCGATG 894
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAenAlaGlySer 140
DB 895 GTGGTTGACTGTGCGGCGCATTAACAACTTCGATTCGATTCGCTTAACCGCGATCG 954
QY 141 LeuGluLysAspLeuGlnGluLysTyArgIluProThrProGlnAlaLeuLeuGluSer 160
DB 955 CTGGAAGAAAGATCTGCAGAAAGATATGCGAACCCACCCGAGCGTGTGCGATGCT 1014
QY 161 AlaMetArgHisValAspHisLeuAspArgLysAsnPheAspGlnPheLysValSerVal 180
DB 1015 GCCATCGCATGTTGATCATCTCGATCGCTCGAAGCTTGATGATCAAGTCAGCGTG 1074
QY 181 LysAlaSerAspValPheLeuAlaValGlnSerTyArgLeuLeuAlaLysGlnIleAsp 200
DB 1075 AAAGCGTTCGACGTCTTCCTGCTGTAGTCTTATCGTTTCTGCGCAAAACGATCGAT 1134
QY 201 GlnProLeuHisLeuGlyLleThrGluAlaGlyValAlaLysSerGlyValAlaLysSer 220
DB 1135 CAGCGCTTCATCTGGGATCACCGAAGCGGTGGCGCGAGCGGCGACGTAAATCC 1194
QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGlyIleGlyAspThrLeuArgValSerLeu 240
DB 1195 GCCATTGGTTAGTCTGCTGCTGTAAGCATTCGGGACACGCTGCGCTATCGCTG 1254
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
DB 1255 GCGGCGGATCCGTCGAAGAGATCAAAAGTGGTTTCGATATTGTAATCGCTGCGTATC 1314

|||||
Db 421 CTGGAAGAAATCGATCGAAGAAAGTATGGCAACCGGACCGCGAGCGGTTCGATCTC 480
QY 161 AAlaMetArgHisValAspHisLeuAspArgLeuAspPheAspGlnPheLeuValSerVal 180
Db 481 GCCATGCGTATGATGTGATCTCGATCGCTGAACTTCATCATCAAGTTCAAAGCAGGTTG 540
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 541 AAGCGCTGACGCTTCCTCGCTGTTGATCTTATGTTTGTGCGGAAAACGATCGAT 600
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyValAlaLysSer 220
Db 601 CAGCCGTCGATCGGGGATCACCGAAGCCGGTGTGCGCGACGCGGCGAGTAAATCC 660
QY 221 AlaIleGlyLeuLeuLysLeuLeuSerGlyGlyIleGlyAspThrLeuAspValSerLeu 240
Db 661 GCCATTGGTTAGTGTGCTGCTGTGAAGGATCGGCAACGCTGCGCATATGCTG 720
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 721 GCGCGCGATCGGTCGAAAGATCAAAAGTCGTTTCGATTTTGAATGCGTCCGATC 780
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnIlePheAspVal 280
Db 781 CGTTCGCGAGGAGTCAACTTCATCGCTGCCGCCACCTGTCGCGTCAAGAAATTTGATGTT 840
QY 281 IleGlyThrValAsnAlaLeuGluGlnArgLeuGluAspIleIleThrProMetAspVal 300
Db 841 ATCGGTAACGGTTAAACGGCTGAGCAACGCTGGAAGATTCATCATCTCGATGAGACGTT 900
QY 301 SerIleIleGlyCysValValAsnGlyProGlyGluAlaLeuValSerThrLeuGlyVal 320
Db 901 TCATATATCGCTGCGCTGGTGAATGGCCACGATGAGGCGCTGTTCTACACTCGGCTC 960
QY 321 ThrGlyValAsnLysSerGlyLeuTyrGlyAspGlyValArgLysAspArgLeuAsp 340
Db 961 ACCGGCGGCAACAAGAAAGCGGCTCTTGAAGATGCGGTGCGCAAGACCGCTCGAC 1020
QY 341 AsnAsnAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
Db 1021 AACCAAGATATGATGATCAACGCTGGAAGCAACGATTCGTGCGAAGCCATCGCTGAC 1080
QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
Db 1081 GAAGCGCTGATTCATTCAGCTTCAGCAGGTTGAAAA 1116
RESULT 5
AX393943 1119 bp DNA linear PAT 23-MAR-2002
LOCUS AX393943
DEFINITION Sequence 3 from Patent WO0212478.
ACCESSION AX393943
VERSION AX393943.1 GI:19701905
KEYWORDS
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS 1
Seeman, A., Campos, N., Rodriguez-Concepcion, M., Rohmer, M.,
Serrano, M., Valentin, H.E., Venkatesh, T.V. and Venkatesh, M.,
Methyl-d-erythritol phosphate pathway genes
Patent: WO 0212478-A 3 14-FEB-2002;
Monsanto Technology LLC (US)
FEATURES
source Location/Qualifiers
1..1119
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
1..1119
/note="unnamed protein product"
/codon_start=1
/cramel_table=11

/protein_id="CAD29065.1"
/db_xref="GI:19701906"
/translation="MNQAPIQRKSTRIVGVNPIGDGAPLAVQSMTRRTTVEAT
VNOIVLRFVQGDIVRVSPVMDAAEFLTIQQUNVPIVADIPFVRLAKVAREGV
DCLRLNPQIGNBERIRNRYVDKARDKNIDIRIKNVAGSLEKIQEYRTPALLS
ARHVDHLRLNFDQFVSVKASDVFLAESTRLAKQIDQPLHITLTAAGRSAGV
KSAIGLGLSLEIGDITLRSVLAADPVEIKGFDILKSLRISRGINIACTCSRO
EFDVIGTVALQRLIEDITPMDVSIIGCVNAGPGALVSTLGTGANKSGLYEDGV
RDRLDNMIDQLEARIKAKASQDEARIDVQVEK"
BASE COUNT 272 a 279 c 317 g 251 t
ORIGIN
Alignment Scores:
Pred. No.: 9,1e-136 Length: 1119
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-09-921-992-50 (1-372) x AX393943 (1-1119)
QY 1 MethIleAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyrValGlyAsnVal 20
Db 1 ATGCATTAACCAAGCTTCATCAACGTAGAAAATCAACGATTAATCGTTGGAAATGTG 60
QY 21 ProIleGlyAspGlyValaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 61 CGATTGGCGATGGTGTCTCCATCGCTGACGTCCATACCAATAGGCGTACAGAC 120
QY 41 ValGluAlaThrValAsnGlnIleLysAlaLeuGluArgValGlyValaAspIleValArg 60
Db 121 GTCGAAGCAACGGTCAATCAATCAAGGCGCTGGAACGGTGGCTGATATCGTCCG 180
QY 61 ValSerValProThrMetAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValAsn 80
Db 181 GATATCCGTACCAACGATGACCGCGCAGAGGCTTCAAACTCATCAACAGCAGTTAAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValaGluTyr 100
Db 241 GTGCGCTGTGTGCTGACATTCATCTTCGATTCGATTCGCTGAAAGTAGCGGAATAC 300
QY 101 GlyValaAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
Db 301 GCGCTGATGCTGTGCTATTAACCTGCGCATATCGTAATGAGAGCGTATTCGATG 360
QY 121 ValValaAspCysAlaArgAspLysAsnIleProIleArgIleGlyValaAsnAlaGlySer 140
Db 361 GTGTTGACTGTGCGCGCATAAACATTCGATCCGATTCGATTAAGCGCGATCG 420
QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThrProGlnAlaLeuLeuGlnSer 160
Db 421 CTGGAAGAAAGATCTCGAAGAAAGATATGCGCAACCAACCGCCGAGCGGTTCGGAATCT 480
QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
Db 481 GCCATGCGTATGATGTGATCTTCATCGCTCCGAACTTCGATCAAGTCAACGACGCTG 540
QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
Db 541 AAGCGCTGACGCTTCCTCGCTGTTGATCTTATGTTTGTGCGGAAAACGATCGAT 600
QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyValAlaArgSerGlyValAlaLysSer 220
Db 601 CAGCCGTCGATCGGGGATCACCGAAGCCGGTGTGCGCGACGCGGCGAGTAAATCC 660
QY 221 AlaIleGlyLeuGlyLeuLeuLysLeuLeuSerGlyGlyIleGlyAspThrLeuAspValSerLeu 240
Db 661 GCCATTGGTTAGTGTGCTGCTGTGAAGGATCGGCAACGCTGCGCATATGCTG 720
QY 241 AlaAlaAspProValGluGluIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 721 GCGCGCGATCGGTCGAAAGATCAAAAGTCGTTTCGATTTTGAATGCGTCCGATC 780

QY 41 ValGluAlaThrValaLsnGlnIleValaLsnGlnuArgValaGlyAlaAspIleValArg 60
 Db 121 GTGAAACCAACGGTCAATCAATCAAGCGCTGGAAGCGCTGGCGTGAATCGCTCGT 180
 QY 61 ValSerValProThrmecAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValaLsn 80
 Db 181 GTATCCCTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 240
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValaAlaGlyTyr 100
 Db 241 GTGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300
 QY 101 GlyValaAspCysLeuArgIleAsnProGlyAsnIleGlyLysGlnGlnuArgIleArgMet 120
 Db 301 GGGCGTCAATGTCTGCTATTAACTTGGCAATATCGGTATGAAGAGCGTATTGGCATG 360
 QY 121 ValValaAspCysAlaArgAspLysAsnIleProIleArgIleGlyValaLsnAlaGlySer 140
 Db 361 GTGCTTACCTGTGGCGCGCATTAACCAATTCGATCCGATTTGGCGTTAACGCCGATCG 420
 QY 141 LeuGluLysAspLeuGlnGlyLysTyrGlyGluProThrProGlnAlaLeuLeuGlySer 160
 Db 421 CTGGAATAAAGATCTGCAAGAAAGTATGGCAACCGACCGCGCGCGTGGTGGATCT 480
 QY 161 AlaMetArgHisValaAspHisLeuAspArgLeuAsnPhelAspGlnPhelLysValSerVal 180
 Db 481 GCCATGGGTATGTATGATCTGATCGCTGATCGCTGATCGCTGATCGCTGATCGCTGAT 540
 QY 181 LysAlaSerAspValaPheLeuAlaValaGlySerTyrArgLeuLeuAlaLysGlnIleAsp 200
 Db 541 AAGCGCTGATCGCTCTCTCGCTGATGATCTTATGCTTTGCTGGCAAAACGATGATGAT 600
 QY 201 GlnProLeuHisLeuGlyIleThrGluAlaGlyLysValaLysSerValaValLysSer 220
 Db 601 CAGCGCTGATCTGCGGATCAACGAAAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
 QY 221 AlaIleGlyLeuGlyLysLeuLeuLeuSerGluGlyIleGlyLysPheLeuArgValSerLeu 240
 Db 661 GCCATTGGTTAAGTGTGCTGCTGTGTAAGGCAATCGCGCAACGCGTGGCGTATCGCTG 720
 QY 241 AlaAlaAspProValaGluGluIleLysValaGlyPheAspIleLeuLysSerLeuArgIle 260
 Db 721 GCGCGCATCCGCTCGAAGACATCAAACTCGCTTCAATTTTGAATCGCTCGCTCGATC 780
 QY 261 ArgSerArgGlyIleAsnPhelAlaCysProThrCysSerArgGlnLysPheAspVal 280
 Db 781 CGTTCGCGAGGATCACTTCATCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 840
 QY 281 IleGlyThrValaLsnAlaLeuGlnuArgLeuGlnuArgIleIleThrProMetAspVal 300
 Db 841 ATCGGTACGGTAAACGGCTGACCAACGCTGGAAGATATCATCATCTCGATGAGCGTT 900
 QY 301 SerIleIleGlyCysValaValaLsnGlyProGlyGlnAlaLeuValaSerThrLeuGlyVal 320
 Db 901 TCGATTATCGCTCGCTGCTGAATGCGCCAGGTAGCGCGCTGTTTCTACCTCGCGCTC 960
 QY 321 ThrGlyLysLeuLysSerGlyLysuArgIleValaLysAspArgLeuAsp 340
 Db 961 ACCGCGGCAACAGAAAGCGCGCTCATATGAATGAGCGCGCAAAAGCGCTGAGAC 1020
 QY 341 AsnAsnAspMetIleAspGlnLeuGlnuArgIleArgAlaLysValaSerGlnLeuAsp 360
 Db 1021 AACACACATATGATGACACGCTGAAAGCAAGCATGTGTGCGAAAGCAAGTCACTGAC 1080
 QY 361 GlnAlaArgArgIleAspValGlnIleValaGlyLys 372
 Db 1081 GAAGCGCGTGAATTAAGCTTCAACAGTTCAAAA 1116
 RESULT 4
 AX050487 1119 bp DNA linear PAT 12-JAN-2001
 LOCUS AX050487
 DEFINITION Sequence 1 from Patent WO0072022.

ACCESSION AX050487
 VERSION AX050487.1 GI:12226691
 KEYWORDS
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
 REFERENCE
 1 Jomaa, H.
 TITLE Use of genes of the deoxy-d-xyulose phosphate biosynthetic pathway for altering the concentration of isoprenoid
 JOURNAL Patent: WO 0072022-A 1 30-NOV-2000;
 Jomaa, Hassan (DE)
 FEATURES
 source
 1..1119 Location/Qualifiers
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /db_xref="taxon:562"
 1..1119
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC21801.1"
 /db_xref="GI:12226692"
 /translation="MHNOAPIORRKRSTRIVGVNPFGDAPIAVQSMNTRRTTVEAT VNOIALERVGADIVRVSPTMDAARAFILIQOVNVPVADHPRIALAKVAEYGV DCLRINPGRNGBERIRMYDDCARDNIPRIGQVNSLEKLOEXKGRPTPALLES AMRHVDHDLRNLNDQPKVSVKASDVFLAVESRLKQIDQPLHGLITEGAKSGAV KSAIGLLLSGSGIDTLRSLADPVEISIKVGFILKSLRISRGINFLACPTCSRO EFDVIGTVALRQRLIEDIITPMQVSIIGCVNPGSALVSTLGVTTGNKKSGLYEDGV RDRIDNNMIDQLEAIRAKASQIDEARIDVQVEK"
 BASE COUNT 272 a 279 c 317 g 251 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 9..1e-136 Length: 1119
 Score: 1866.00 Matches: 372
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: Gaps: 0
 US-09-921-992-50 (1-372) x AX050487 (1-1119)
 QY 1 MethiAsnGlnAlaProIleGlnArgGlySerThrArgIleTyrValaGlyAsnVal 20
 Db 1 ATGCATTAACCGAGCTCCAAATTCACGTAGAAATCAACGATTTTAACGTGGAAATGTG 60
 QY 21 ProIleGlyAspGlyAlaProIleAlaValaGlnSerMetThrAsnThrArgThrAsp 40
 Db 61 CCGATTGGCGATGTCTCCCATCGCGCTGACATCCATGACCAATACGCTAGACACGAC 120
 QY 41 ValGluAlaThrValaLsnGlnIleLysAlaLeuGlnuArgValaGlyAlaAspIleValArg 60
 Db 121 GTGAAACCAACGGTCAATCAATCAAGCGCTGGAAGCGCTGGCGCTGATCGCTCGT 180
 QY 61 ValSerValProThrmecAspAlaAlaGluAlaPheLysLeuIleLysGlnGlnValaLsn 80
 Db 181 GTATCCCTACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 240
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuLysValaAlaGlyTyr 100
 Db 241 GTGCGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 300
 QY 101 GlyValaAspCysLeuArgIleAsnProGlyAsnIleGlyLysGlnGlnuArgIleArgMet 120
 Db 301 GGGCGTCAATGTCTGCTATTAACTTGGCAATATCGGTATGAAGAGCGTATTGGCATG 360
 QY 121 ValValaAspCysAlaArgAspLysAsnIleProIleArgIleGlyValaLsnAlaGlySer 140
 Db 361 GTGCTTACCTGTGGCGCGCATTAACCAATTCGATCCGATTTGGCGTTAACGCCGATCG 420
 QY 141 LeuGluLysAspLeuGlnGlyLysTyrGlyGluProThrProGlnAlaLeuLeuGlySer 160

VERSION AX036302.1 GI:11225912
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS
TITLE Novel method for identifying antibacterial compounds
JOURNAL Patent: EP 1043403-A 29 11-OCT-2000;
GPC AG GENOME PHARMACEUTICALS (DE)
FEATURES
source 1. 1119
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
BASE COUNT 272 a 279 c 317 g 251 t
ORIGIN
Alignment Scores:
Pred. No.: 9,1e-136 Length: 1119
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-09-921-992-50 (1-372) x AX036302 (1-1119)
QY 1 MethHAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyValGlyAsnVal 20
Db 1 ATGCATACCAAGCGCTCCATTCAACGTAAAGAAATCAACAGTATTACGTGGGAAGTG 60
QY 21 ProIleGlyAsnGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 61 CCATATGGCGATGGTGTCTCCATCGCGTACAGTCCATACCAATACGATACGACAGAC 120
QY 41 ValGlnAlaThrValAsnGlnIleTyValAlaGlnArgValGlyAlaAspIleValArg 60
Db 121 GTGCAAGCAACGGCTCAATCAATCAAGCCCTGGAACCGCTGGCGCTGATTCGTCCT 180
QY 61 ValSerValProThrMetAspAlaAlaGlnAlaPheLysLeuIleLysGlnGlnValAsn 80
Db 181 GTATCCGTCACGACGATGACGCGGCGGAGAGCGTTCAACTCAATCAACAGCAGTTTAC 240
QY 81 ValProLeuValAlaAspIleHisPheAspTyArgIleAlaLeuLysValAlaGlnTyr 100
Db 241 GTCCCGGTGGTGGCTGCATCCACTTCGACTATCGCATTCGCTGAAAGTAGCGGATAC 300
QY 101 G1ValAspCysLeuArgIleAsnProG1ValAsnIleG1ValAsnGlnG1ValArgMet 120
Db 301 GGGCGCATTTGTCTGGATTTAAACCTGCGCAATTCGATTAAGAGCGTATTCGATG 360
QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleG1ValAsnAlaGlySer 140
Db 361 GTGGTTGACTGTGGCGCGCATTAACATTCGATTCGATTCGATTCGATTCGATTCG 420
QY 141 LeuGlnLysAspLeuGlnGlnLysTyArgLysGlnProThrProGlnAlaLeuLeuGlnSer 160
Db 421 CTGCAAAAGATCTGCAGAAAGAAAGTATGGCAACCGACGCGAGGGTGTCTGATCTC 480
QY 161 AlMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheLysValSerVal 180
Db 481 GCCATGGGTATGTTGATCTCTCGATCGCTGATTCGATTCGATTCGATTCGATTCG 540
QY 181 LysAlaSerAspValPheLeuAlaValGlnSerTyArgLeuLeuAlaLysGlnIleAsp 200
Db 541 AAGCGCTGACGCTCTTCCTCGCTGTGATCTTATCTTATCTTATCTTATCTTATCTT 600
QY 201 GlnProLeuHisLeuGlnIleThrGlnAlaGlyGlyAlaArgSerGlyAlaValLysSer 220
Db 601 CAGCCGTTCATCTGGGATCACCGAAGCGGTGTGCGCGACGGGGCGAGTAAATCC 660

QY 221 AlaIleGlyLeuGlnLeuLeuSerGlnGlyIleGlyAspThrLeuArgValSerLeu 240
Db 661 GCCATTGGTTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
QY 241 AlaAlaAspProValGlnGlnIleLysValGlyPheAspIleLeuLysSerLeuArgIle 260
Db 721 GCGCCGATCCCGTCCAGAGATCAAGATCGATTTGATTTGAAATCGCTGCGTATC 780
QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThrCysSerArgGlnGlnPheAspVal 280
Db 781 CATTCCGAGGATCAACTCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
QY 281 IleGlyThrValAsnAlaLeuGlnGlnArgLeuGlnAspIleIleThrProMetAspVal 300
Db 841 ATCGGTACGGTTAACCGCTGAGCAACCGCTGGAAGATATCATCTCGATGGAAGCTT 900
QY 301 SerIleIleGlyCysValValAlaAsnGlyProGlyGlnAlaLeuValSerThrLeuGlyVal 320
Db 901 TCGATTATCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
QY 321 ThrGlyValAsnLysSerGlyLeuTyArgLysAspGlyValArgLysAspArgLeuAsp 340
Db 961 ACCGCGCGCAACAGAAAGCGGCTCTTATGAAGATGCGCTGCGCAAGACCGTCTGGAC 1020
QY 341 AsnAsnAspMetIleAspGlnLeuGlnAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
Db 1021 AACCAACGATATGATTCACCAAGCTGGAAGCAGCATTCGTGCGAAGCCAGTCACTGGAC 1080
QY 361 GlnAlaArgArgIleAspValGlnGlnValGlnLys 372
Db 1081 GAAGCGCGTCAATGACGTTGACGAGTGGAAAA 1116
RESULT 3
AX038912 1119 bp DNA linear PAT 16-NOV-2000
LOCUS Sequence 29 from Patent WO061793.
DEFINITION AX038912
ACCESSION AX038912
VERSION AX038912.1 GI:11228221
KEYWORDS
SOURCE Escherichia coli
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Loferer H. and Jacobi A.
TITLE Novel method for identifying antibacterial compounds
JOURNAL Patent: WO 0061793-A 29 19-OCT-2000;
GPC BIOTECH AG (DE) ; LOFERER HANNES (DE) ; JACOBI ALEXANDER (DE)
FEATURES
source 1. 1119
/organism="Escherichia coli"
/mol_type="genomic DNA"
/db_xref="taxon:562"
BASE COUNT 272 a 279 c 317 g 251 t
ORIGIN
Alignment Scores:
Pred. No.: 9,1e-136 Length: 1119
Score: 1866.00 Matches: 372
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
Gaps: 0
US-09-921-992-50 (1-372) x AX038912 (1-1119)
QY 1 MethHAsnGlnAlaProIleGlnArgArgLysSerThrArgIleTyValGlyAsnVal 20
Db 1 ATGCATACCAAGCGCTCCATTCAACGTAAAGAAATCAACAGTATTACGTGGGAATG 60
QY 21 ProIleGlyAsnGlyAlaProIleAlaValGlnSerMetThrAsnThrArgThrAsp 40
Db 61 CCATATGGCGATGGTGTCTCCATCGCGTACAGTCCATACCAATACGATACGACAGAC 120

AY033515
 LOCUS AY033515 1119 bp DNA linear BCT 26-DEC-2001
 DEFINITION Escherichia coli 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (ispG) gene, complete cds.
 ACCESSION AY033515
 VERSION AY033515.1 GI:17978527
 KEYWORDS
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
 REFERENCES
 Hecht, S., Eisenreich, W., Adam, P., Amslinger, S., Kis, K., Bacher, A., Arigoni, D., and Rohdich, F.
 Studies on the nonmevalonate pathway to terpenes: The role of the Gcpe (ispG) protein
 Proc. Natl. Acad. Sci. U.S.A. 98 (26), 14837-14842 (2001)
 11752431
 2 (bases 1 to 1119)
 Rohdich, F., Hecht, S., Adam, P., Bacher, A. and Eisenreich, W.
 Direct Submision
 Submitted (27-APR-2001) Institut fuer Organische Chemie und Biochemie, Lehrstuhl III, Technische Universitaet Muenchen, Lichenberg Strasse 4, Garching D-85747, Germany
 FEATURES
 source
 1. 1119
 location/Qualifiers
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /db_xref="taxon:562"
 1. 1119
 /gene="ispG"
 1. 1119
 /gene="ispG"
 /function="converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate, last but one step of the mevalonate-independent pathway for isoprenoid biosynthesis"
 /note="ispG"
 /codon_start=1
 /transl_table=11
 /product="1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase"
 /protein_id="AAK53460.1"
 /db_xref="GI:17978528"
 /translation="MHQAPIORRKSRIYGVNPIGDGAPIAVQSMTRTTRTDEAT VNOIKALERVADIVRVSVPTMDAAFAFKLIKQOVNPLVADHFDRIALKYAEV DCLRINGNIGNEERIRMVVDCAKDKNIPRIGVAGSLKEDLOEKYGEPTPOLLES AMRVHDLDELNPQPKVSYSKASDVPLAVESYRLAQIDOPHLGITTEGARSQV KSATGLULSEGIQDTRVSLADPVEETIVGDIKSLRIRKGINFIACPTCSKQ EFDVIGTVNALEQLEDIITPMDSIIIGCVNNGEALVSTLGVGTGNKKSGLYEDGV RKRLDNNMDMDLEARIKAKASQLEARRIDVQOVER"
 BASE COUNT 272 a 280 c 315 g 252 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 9.1e-136 Length: 1119
 Score: 1866.00 Matches: 372
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 1 Gaps: 0
 US-09-921-992-50 (1-372) x AY033515 (1-1119)
 QY 1 MethisanginalaProileGlnArgArgThraGlyIleYrValGlyAsnVal 20
 DB 1 ATGATATACCGGCTCCCAATTCACGTAGAAATCAACACGATATTACGTTGGAAAGTG 60
 QY 21 ProileGlyAspGlyAlaProileAlaValGlnSerMetThraSerThraArgThrThraAsp 40
 DB 61 CCGATTGGCGATGGTCCCATCGCGGTACAGTCAATGACCAATACGCGGTACACAGAT 120

QY 41 ValGluAlaThraValAsnGlnIleuysAlaLeuGluArgValGlyValAspIleValArg 60
 DB 121 GTGAAGCAACGGTCAATCAATCAAGCGCTGGAAAGCGTTGGCCCTGATATCGTCCT 180
 QY 61 ValSerValProThraMetAspAlaAlaGluAlaPheLeuLeuIleuysGlnGlnValAsn 80
 DB 181 GTCCCGTACCGACGATGAGCGCGGCAAGAGCTTCAAACTATCAAAACGACGATTAC 240
 QY 81 ValProLeuValAlaAspIleHisPheAspTyrArgIleAlaLeuysValAlaGluTyr 100
 DB 241 GTGCCCTGGTGGCTACATCCACTTCGACTACGATTCGCTGAAAGTAGCGGAATAC 300
 QY 101 GlyValAspCysLeuArgIleAsnProGlyAsnIleGlyAsnGluGluArgIleArgMet 120
 DB 301 GSGCTGATGCTGCTGATTAACCTCGCAATATCGTAAGAGAGCGATTCGCATG 360
 QY 121 ValValAspCysAlaArgAspLysAsnIleProIleArgIleGlyValAsnAlaGlySer 140
 DB 361 GTGGTTGACGTGCGCGCGATAAACATTCCGATTCGATTGGCGTTAACCGCGATCG 420
 QY 141 LeuGluLysAspLeuGlnGluLysTyrGlyGluProThraProGlnAlaLeuLeuGluSer 160
 DB 421 CTGGAATAAGATCTGCAAGAAAGTATGGCAACCGACCGCGCGCTGCTGGAATCC 480
 QY 161 AlaMetArgHisValAspHisLeuAspArgLeuAsnPheAspGlnPheValSerVal 180
 DB 481 GCCATCGCATGTTATATCTCGATCGCTGAACCTCGATCGATTCGATCAAGTCAGCGTG 540
 QY 181 LysAlaSerAspValPheLeuAlaValGluSerTyrArgLeuLeuAlaLysGlnIleAsp 200
 DB 541 AAGCGTCTGACCTCTCCCTCGCTGTGATCTTACGTTTGTGCAAAACAAATCGAT 600
 QY 201 GlnProLeuHisLeuGlyIleThraGluAlaGlyValAlaArgSerGlyValAlaLysSer 220
 DB 601 CACCGCTGATCTGGGATCAACGGAAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660
 QY 221 AlaIleGlyLeuGlyLeuLeuLeuSerGluGlyIleGlyAspThrLeuArgValSerLeu 240
 DB 661 GCCATTGTTAGTCTGCTGCTGCTGCTGCAAGGATCGGACACGCTGCGCTGATCATCTG 720
 QY 241 AlaAlaAspProValGluGluIleuysValGlyPheAspIleLeuysSerLeuArgIle 260
 DB 721 CGCGCGCATCGCTCGAAGATCAAGATCGATTGATTTGAAATCGCTTCGATC 780
 QY 261 ArgSerArgGlyIleAsnPheIleAlaCysProThraCysSerArgGlnGluPheAspVal 280
 DB 781 CGTTCCGAGAGTATCAACTTCATCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 840
 QY 281 IleGlyThraValAsnAlaLeuGluGlnArgLeuGluAspIleIleThraProMetAspVal 300
 DB 841 ATCGGTACAGTTAACCGCTGGGCAACGCTGGAAGATATCACTCCGATGAGCGTT 900
 QY 301 SerIleIleGlyCysValValAsnGlyProGlyGlyAlaLeuValSerThraLeuGlyVal 320
 DB 901 TCGATTATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
 QY 321 ThrGlyGlyAsnLysLysSerGlyLeuTyrGluAspGlyValAlaArgLysAspArgLeuAsp 340
 DB 961 ACCGGCGGCAACAAAGAGCGCTCTCTATGAAGATGCGCTGCGCAAGCGCTGCGAC 1020
 QY 341 AsnAspAspMetIleAspGlnLeuGluAlaArgIleArgAlaLysAlaSerGlnLeuAsp 360
 DB 1021 AACACGATATGATCGACGACGCTGAGGACGCGATTCGTCGAAAGCCATCGACTGAGAC 1080
 QY 361 GluAlaArgArgIleAspValGlnGlnValGluLys 372
 DB 1081 GAAGCGGTGGAATTCGCTTACGAGGTGAAAAA 1116
 RESULT 2
 LOCUS AX036302 1119 bp DNA linear PAT 16-NOV-2000
 DEFINITION Sequence 29 from Patent EP1043403.
 ACCESSION AX036302

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: November 23, 2003, 17:17:44 ; Search time 4608 Seconds
(without alignments)
3302.600 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866
Sequence: 1 MHNQAPIQRKRSRIYGVN.....RAKASQDEARIDVQVER 372

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 segs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xh
-Q/cgn2_1/USFTO.spool/US0921992/rnuc_21112003_162546_18406/app_query.fasta_1.519
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45
-DOCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pic -NORM=exc -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0921992.OCN 1.1 3508.0runat_21112003_162546_18406 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEOUTRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_ey: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *

29: em_vi: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_hcg_mus: *
34: em_hcg_pln: *
35: em_hcg_rtd: *
36: em_hcg_mam: *
37: em_hcg_vrt: *
38: em_sy: *
39: em_hcg_hum: *
40: em_hcg_mus: *
41: em_hcg_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1866	100.0	1119	1 AY033515	AY033515 Escherich
2	1866	100.0	1119	6 AX036302	AX036302 Sequence
3	1866	100.0	1119	6 AX038912	AX038912 Sequence
4	1866	100.0	1119	6 AX050487	AX050487 Sequence
5	1866	100.0	1119	6 AX393943	AX393943 Sequence
6	1866	100.0	1637	1 ECGEPE	E64451 E.coli GGPB
7	1866	100.0	11521	1 AE005481	AE005481 Escherich
8	1866	100.0	13176	1 AE000338	AE000338 Escherich
9	1866	100.0	296827	1 AP002561	AP002561 Escherich
10	1866	100.0	300099	1 AE016764	AE016764 Escherich
11	1855	99.4	10225	1 AE015271	AE015271 Shigella
12	1855	99.4	290380	1 AE016987	AE016987 Shigella
13	1840	98.6	23647	1 AE008814	AE008814 Salmonella
14	1831	98.1	145050	1 AL627275	AL627275 Salmonella
15	1831	98.1	300411	1 AE016835	AE016835 Salmonella
16	1691	90.6	10101	6 AR027993	AR027993 Sequence
17	1691	90.6	1010	6 AR208328	AR208328 Sequence
18	1691	90.6	1010	6 BD084908	BD084908 Methods f
19	1645.5	88.2	301235	1 AP005075	AP005075 Vibrio pa
20	1631.5	87.4	301442	1 AE016798	AE016798 Vibrio vu
21	1624.5	87.1	1345	6 PSU67933	PSU67933 Providencia
22	1624.5	87.1	1345	6 AR027991	AR027991 Sequence
23	1624.5	87.1	1345	6 AR208326	AR208326 Sequence
24	1624.5	87.1	1345	6 BD084906	BD084906 Methods f
25	1622.5	87.0	11161	1 AE013738	AE013738 Yersinia
26	1622.5	87.0	204050	1 AJ414154	AJ414154 Yersinia
27	1621.5	86.9	11824	1 AE004161	AE004161 Vibrio ch
28	1603.5	85.9	10507	1 U32721	U32721 Haemophilus
29	1603.5	85.9	110000	6 AR274513_03	Continuation (4 of
30	1599.5	85.4	11268	1 AE006237	AE006237 Pasteurel
31	1593.5	85.4	1079	6 AR027992	AR027992 Sequence
32	1593.5	85.4	1079	6 AR208327	AR208327 Sequence
33	1593.5	85.4	1079	6 BD084907	BD084907 Methods f
34	1584.5	84.9	10607	1 AE015769	AE015769 Shewanell
35	1372.5	73.6	17148	1 AE004798	AE004798 Pseudomon
36	1353	72.5	886	6 AR027995	AR027995 Sequence
37	1353	72.5	886	6 AR208330	AR208330 Sequence
38	1353	72.5	886	6 BD084910	BD084910 Methods f
39	1341.5	71.9	300661	1 AE016777	AE016777 Pseudomon
40	1335.5	71.6	311249	1 AE016860	AE016860 Pseudomon
41	1285.5	68.9	9608	1 AE014104	AE014104 Buchnera
42	1282	68.7	110000	6 BD061520_3	Continuation (4 of
43	1282	68.7	347550	1 AP001118	AP001118 Buchnera
44	1226.5	65.7	349174	1 AB063522	AB063522 Wigglestwo
45	1186	63.6	4775	1 D90881	D90881 E.coli geno

ALIGNMENTS

RESULT 1

; SEQ ID NO 8947
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8947

Query Match 5.3%; Score 98.5; DB 15; Length 365;
Best Local Similarity 20.9%; Pred. No. 0.24;
Matches 77; Conservative 54; Mismatches 138; Indels 99; Gaps 19;

```

QY 18 GNPVIGGAPVIAQSMNTRTDVEA-----TVNQIKALERVGAD----- 57
Db 28 GSAPVGGGAVLHGTGEGAGVSVFARPGAYSTLRDLVDRFRAELVGADPFQREWLHR 87
QY 58 -----IVRSVPPTMDAAE-----AFKLIXO-QVNVPLVADHFDYRIA-- 94
Db 88 VMELDRHHELPPLTGLGIDIALMDLAGRYGEPVWRLLGGFRRAIPAYASTFSSVAEF 147
QY 95 LKVAEYGVDCRLINPGNIGNEERIRMVVDCARDKNIPIRIGVNAAGSLEKDLQEKY---G 150
Db 148 LDVAD---QCLALGYRGI---KLHAWGDARDDELCLALRDHVG---PDVPLMYDGSAG 197
QY 151 EPTPQAL-LESAMRHVDLRLNFDQKVSVDVFLAVESYRLAKQIDQPLHLGITE 209
Db 198 FDLPDALRLGRALSEADYL-----WYEEPIRE---FSISAYQRLAEAVDVPLVAETS 247
QY 210 AGG-----ARSGAVKSAIGLGLLSEGIQDTRLRVSLADPVEEIKVGFILKSLRIRS 262
Db 248 DGAHMAADPIRAGAAITFGVRAGTTLRGILTGAMKTHLADAFR-----LRAEV 296
QY 263 RGI---NFIACTPSRQEF--DVIGTVNALQRLIEDIITPMVDSIIGCVVNGPGEALVST 317
Db 297 HGSDDIPVHHLQMAISNTTYESLVTSVNVVRERHVD-----DQGLVHAPA-GPGIALPLD 350
QY 318 LGVTGK 325
Db 351 FGX--GNE 356

```

Search completed: November 23, 2003, 17:22:06
Job time : 50 secs

;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 5635
;; LENGTH: 2368
;; TYPE: PRF
;; ORGANISM: Staphylococcus aureus
US-09-815-242-5635

Query Match 5.3%; Score 99; DB 9; Length 2368;
Best Local Similarity 20.8%; Pred. No. 4.1;
Matches 89; Conservative 58; Mismatches 155; Indels 126; Gaps 19;

QY 3 NOAPIQ-RRKSTRIYGVNPIGDS-----APIAQSMNTTRTVE--- 42
DB 1532 NQAKTNIDSSIDBYDNA-VKEGAKINAVKTFSEYKDALAKIEAAYNAKVTENDNSN 1590
QY 43 -ATVNOIKALERVGADIVRSVPTMDAAEAFKLIKQOVNVPVADIHFDYRIALKVAEYG 101
DB 1591 ASTSSIEAEKQKLAELKQTADQNVQATS---KODIEVQIHNDLD----- 1633
QY 102 VDCLRINPGNIGNEERIRNVDC--ARDKNIPIRIGVNAAGSLEKDLQEKYGEPTQALL 158
DB 1634 ---NINDYTIPTGKESATTDLYAYADQKNNISADTNATQDEKQALIKQVDQNVQTAL 1689
QY 159 ESAMRHVDH-----LDRLNFD-----QFKSVKASDVFLAV 189
DB 1690 ESINNGVDNDVDALTOGKAALDAIQVDATYKPRANQVIDAKAEETKESIDSDQLTAE 1749
QY 190 ESYRLA--KOIDPRLHIGTEAGARGAVKSAIGLLSEGIG-DTLRVSILADPVE 246
DB 1750 EKTEALAMIKQITDQAKQGITDA--TTTAEVEKAKAQGLEAFNDIQIDSTEKQKA---IE 1804
QY 247 EIKVGFDIKSLIRSRGINFIACPTCSQEQEDVIGTVNALSQRLEDITTPMDVSIIGCV 306
DB 1805 ELETALDQIEA-----GVNVDADATTEKE-----AFTNALIEDISKATEDISDQT 1850
QY 307 VNGPGEALVSTLGVGTGKNSGLYEDGVKRDRLNNMDIDQLEA-RIR--AKASQDEAR 363
DB 1851 TN---AEIATV-----KNSALEQLKAKORINPVVKKNALEAIR 1884
QY 364 RIDVOQVE 371
DB 1885 EVVNNQIE 1892

RESULT 14
US-09-815-242-12389
; Sequence 12389, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haseelbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 12389
;; LENGTH: 2368
;; TYPE: PRF
;; ORGANISM: Staphylococcus aureus
US-09-815-242-12389

Query Match 5.3%; Score 99; DB 9; Length 2368;
Best Local Similarity 20.8%; Pred. No. 4.1;
Matches 89; Conservative 58; Mismatches 155; Indels 126; Gaps 19;

QY 3 NOAPIQ-RRKSTRIYGVNPIGDS-----APIAQSMNTTRTVE--- 42
DB 1532 NQAKTNIDSSIDBYDNA-VKEGAKINAVKTFSEYKDALAKIEAAYNAKVTENDNSN 1590
QY 43 -ATVNOIKALERVGADIVRSVPTMDAAEAFKLIKQOVNVPVADIHFDYRIALKVAEYG 101
DB 1591 ASTSSIEAEKQKLAELKQTADQNVQATS---KODIEVQIHNDLD----- 1633
QY 102 VDCLRINPGNIGNEERIRNVDC--ARDKNIPIRIGVNAAGSLEKDLQEKYGEPTQALL 158
DB 1634 ---NINDYTIPTGKESATTDLYAYADQKNNISADTNATQDEKQALIKQVDQNVQTAL 1689
QY 159 ESAMRHVDH-----LDRLNFD-----QFKSVKASDVFLAV 189
DB 1690 ESINNGVDNDVDALTOGKAALDAIQVDATYKPRANQVIDAKAEETKESIDSDQLTAE 1749
QY 190 ESYRLA--KOIDPRLHIGTEAGARGAVKSAIGLLSEGIG-DTLRVSILADPVE 246
DB 1750 EKTEALAMIKQITDQAKQGITDA--TTTAEVEKAKAQGLEAFNDIQIDSTEKQKA---IE 1804
QY 247 EIKVGFDIKSLIRSRGINFIACPTCSQEQEDVIGTVNALSQRLEDITTPMDVSIIGCV 306
DB 1805 ELETALDQIEA-----GVNVDADATTEKE-----AFTNALIEDISKATEDISDQT 1850
QY 307 VNGPGEALVSTLGVGTGKNSGLYEDGVKRDRLNNMDIDQLEA-RIR--AKASQDEAR 363
DB 1851 TN---AEIATV-----KNSALEQLKAKORINPVVKKNALEAIR 1884
QY 364 RIDVOQVE 371
DB 1885 EVVNNQIE 1892

RESULT 15
US-10-156-761-8947
; Sequence 8947, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109

QY 330 -----YEDGVKRDRLNNNDMDIDLEAIRAKASQDLEARRIDV 367
Db 517 SVEADAGMGEAGISDDEDTQATEVEKEVVSVEADAGMGEEL--VDV 561
RESULT 11
US-10-306-905-11
; Sequence 11, Application US/10306905
; Publication No. US20030167513A1
; GENERAL INFORMATION:
; APPLICANT: MOURAD, GEORGE S.
; APPLICANT: JUNK, DAMIAN J.
; TITLE OF INVENTION: SELECTION AND USE OF ISOPROPYLMALATE SYNTHASE (IPMS)
; TITLE OF INVENTION: MUTANTS DESSENSITIZED IN L-LEUCINE NEGATIVE FEEDBACK
; FILE REFERENCE: 3220/93981
; CURRENT APPLICATION NUMBER: US/10/306,905
; CURRENT FILING DATE: 2002-11-27
; PRIOR APPLICATION NUMBER: 60/339,895
; PRIOR FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-306-905-11

Query Match 5.4%; Score 101; DB 12; Length 523;
Best Local Similarity 20.3%; Pred. No. 0.24;
Matches 83; Conservative 62; Mismatches 142; Indels 122; Gaps 18;

QY 12 STRIVGNVPIGDPAPIAVQSMTRTTDYATATNOIKALERVGADIVRSVPTMDA-- 69
Db 2 SQOVIIFDTLRDEQALQASLS-----VEKQLIALALERGVGDVWEVGFVSSGDF 55
QY 70 EAFKLIRQOVN---VPLVADIFPDYRIA--LKVAEYGVDCIRI-----NPGNIGNE- 115
Db 56 ESVGTLARQVNSVCLARCEKDIVAASLVAE---AFRIHFIAISPHIATKL 111
QY 116 -----ERIRMVDCAR-----DKNIPRIGVNASGLEKDIQEK 148
Db 112 RSTLDEVIERAIYVWKARNTDDEFEESCEDAGRTPIADLARVVEAIIINGATTINIPDT 171
QY 149 YGEPTP---QALLSARHVDHDLNLPDQKVSVKA--SDVFLAVESYRLAKQIDPRLH 204
Db 172 VGYTMRPEFAGTISGLYERVP---NIDKAIISVHTHDDGLAVGN-SLAA-----VH 219
QY 205 LGITEAGAGASGAVKSAIGLLISEGIDTLRVSLADPVEEIKVGFILKSLIRSRG 264
Db 220 AGARQVSGAMN-----GIGERAGNCSLEVIAMTKVKKDL----- 255
QY 265 INFIACPTCSRQEPDVIGTVNALE-QRLIEDIITPMDVSIIGCVNGEGEALVSTLGYTG 323
Db 256 -----NVHTAIHQEIMWRTSQVLSQ-----ICNMPIDIPANKAIVGSGAF 293
QY 324 NKKSGLYEDGVKRDRLNNNDM-----IDQLEAIRAKASQDLEARRID 366
Db 294 AHSSGHIDQVLRKRENYEIMTPESIGILNQIQLMITSRGAALVXHRMD 342

RESULT 12
US-10-306-905-15
; Sequence 15, Application US/10306905
; Publication No. US20030167513A1
; GENERAL INFORMATION:
; APPLICANT: MOURAD, GEORGE S.
; APPLICANT: JUNK, DAMIAN J.
; TITLE OF INVENTION: SELECTION AND USE OF ISOPROPYLMALATE SYNTHASE (IPMS)
; TITLE OF INVENTION: MUTANTS DESSENSITIZED IN L-LEUCINE NEGATIVE FEEDBACK
; FILE REFERENCE: 3220/93981
; CURRENT APPLICATION NUMBER: US/10/306,905

; CURRENT FILING DATE: 2002-11-27
; PRIOR APPLICATION NUMBER: 60/339,895
; PRIOR FILING DATE: 2001-11-30
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 573
; TYPE: PRT
; ORGANISM: Streptomyces coelicolor
US-10-306-905-15

Query Match 5.3%; Score 99.5; DB 12; Length 573;
Best Local Similarity 20.0%; Pred. No. 0.39;
Matches 82; Conservative 57; Mismatches 131; Indels 141; Gaps 20;

QY 18 GNPVIDGAPIAVQSMTRTTDYATATNOIKALERVGADIVRSVPTMD----- 67
Db 98 GAIP--DVTITV--LTQAREDLIERTVESLKGARATVHLFNATAPVRRVFRGSRDD 153
QY 68 ---AAEAFKLIRQOVNVPVADIFPDYR-----LAKVAEYGVDCIRINPGN 111
Db 154 IKQIAYDGRVLWVEIAEKLGPETFQYQSPRIFTDTLDPALBVCNAVMDTYQCPG- 212
QY 112 IGNEERIRMVVDCARDKNIPRIGVNASGLEKDIQEKYGEPTPQALLSARHV-----D 166
Db 213 -----REII-----LNPATYERSTPSTHADRFEMGRNL-----SREHVCLSVHP 254
QY 167 HDRLNFPDQKVSVKASDVFLAVESYRLAKQIDQPLHIGITEAGAGASGAVSAIGLL 226
Db 255 HNR-----GTVAABAEALMAGADRI-----BGCLFGQGERGNV-DLVTLGM 297
QY 227 -LISEG-----IGDTRLSLAADPVEEIKVGFDI 254
Db 298 NLFSGVDQIDPSDIDELRTWECYQOMEVHRHRYVEDLVYTSQSHQDAIKKGFDA 357
QY 255 LKSLIRSRGINFIACPTCSRQEPDVIGTVNALEQRLEDI-ITPMDSIIGCVNGEGEA 313
Db 358 MEA-DAARGV-----TVDDIEMAVPIPIIDPKDV-----GRS 389
QY 314 LVSTLGYTGANKKSGI-YEDGVKRDRLNNNDMDIDLEAIRAKASQDLEARR 363
Db 390 YEAVIRVNSQSGKGIAVY-----LKNDHSID-LPRRMQIEFSKLIOAK 432

RESULT 13
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

```

QY 282 -----GTVAALQRL--EDITPMDSVIIGCVNPGPALVSTGVTGAGKKKSGLYEDG 333
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 849 IKGDAGSGVALESSLQLQDVGSEVDIRLH---RGVAGVYESDINLATG----- 895
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 334 VRKDRLDNNMIDQLFARIRAKASQLDEARRDIYQ 368
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 896 -----SDRITVIGFVNRAGRAQDAEREGVDVR 923
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 9

```

US-09-159-469-50
Sequence 50, Application US/09159469
Patent No. US20020064535A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
TITLE OF INVENTION: THERAPY OF EHRLICHIA INFECTION
NUMBER OF SEQUENCES: 73
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Avenue, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/159,469
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/106,582
FILING DATE: 29-JUN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Markl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.439C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-682-6031
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 578 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-159-469-50

```

Query Match	5.7%;	Score 105.5;	DB 9;	Length 578;
Best Local Similarity	19.5%;	Pred. No. 0.095;		
Matches 91;	Conservative 77;	Mismatches 176;	Indels 123;	Gaps 21.

```

QY 7 IQRRKSRIRYGNP-----IGDA-----PIAVOSMTNT-PIVETVNO- 47
Db 112 VOEBEGTGMVLIINAPKAVRPFKIEKSAEEPQTUDPSVVESATSGSVDTQEEQIDQ 171
QY 48 IKALIERGAD-----IVRVSPFMDAAEAF--KLIIQVAVP 82
Db 172 APAIEEVETEEOEVILEEGTIDLEQPAQVPAVAEALPGVEAEAAIVPSLEENKLOEV 231
QY 83 LVADIHIDYRALKY--AEYGVDCLRINPENGIGNEEIRKMYVDCARKNPIRIGVANQ 139
Db 232 VVAPEAOOLESAPEVASAPQCESTVLGVAEIDLKSEVSVEANADVAQKE-----VISG 284
QY 140 SLEQDLOEKY-GEFTPOLLE--SAMRVVDHLDRINPQKYSV----- 180

```

```

Db      285  QQGEQIAALBLSGTAPVBEKVETEVLLKEDTLIDLEQPAQVPVAAEALPGVEAAEALV 344
QY      181  -----KASDYFLAVESYRL-LAKOIDOPLH-----LGTTEGARGSAVKSAIGL----- 224
Db      345  PSLEENKIQEVEVVAPEAOQLSAPESVSAPEQESTVLGVT-----GLKBEVSVEADA 398
QY      225  GLLISEGIGDTLRVSLADPVEEIKVGD-----ILKSRLRSRGINFIACTPSRQ 276
Db      399  GMQGEAGISD--QETQATEBEVKESVSVEKTEBEVEILBESLTIDLEBPVQVPVAAEA 456
QY      277  EFDYIGTVNAL-----EQRLDITTPMDVSIIGCV--VNGPEALVSTLGTGNGKKSGL 329
Db      457  ELPGVEAAEALVPSLEENKIQEVEVVAPEAOQLSAPESVSAPEQESTVLGVTGDLKSEV 516
QY      330  -----YEDGYAKDRLDNNNDIMDLEAARIRARASQLDEARRIDV 367
Db      517  SVEADAGMQGEAGISDQETQATEBEVKESVSVEALDAGMQQEL--VDV 561

```

RESULT 10

```

US-09-798-042-50
; Sequence 50, Application US/09798042
; Patent No. US2002006843A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Iodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; AND TREATMENT OF EHRLICHIA INFECTION
; FILE REFERENCE: 210121.439C7
; CURRENT APPLICATION NUMBER: US/09/798,042
; CURRENT FILING DATE: 2001-03-02
; NUMBER OF SEQ. ID NOS: 108
; SOFTWARE: Fastseq for Windows Version 3.0
; SEQ ID NO 50
; LENGTH: 578
; TYPE: PR1
; ORGANISM: Ehrlichia sp.
US-09-798-042-50

```

Query Match	5.7%;	Score 105.5;	DB 9;	Length 578;
Best Local Similarity	19.5%;	Pred. No. 0.095;		
Matches	91;	Conservative	77;	Mismatches 176;
				Indels 123;
				Gaps 21;

Qy	7	IQRKRSRIYVGNP-----IGDA-----PIAVOSMTNRT-PIVEATVNG-	47
Db	112	VOEEEGTGMYLINAPKAVRFFKIEKSAABEPTQDSVVEASAGSGVDYQEEBIDE	171
Qy	48	IKALERVGAD-----IVRVSPPTMDAAEAF--KLIKOVNVP	82
Db	172	APALIEVETEGEVLIEBGLTIDLEQPAQVPVAAEALPGVEAAEALVPSLEENKLG	231
Qy	83	LVADIHEDYRIALKV--AEYGVDCLRINPGNIGNEERIRNVDCARDKNIPRIGVNG	139
Db	232	VVAPEAQOLIESAPVSPAPQPESTVLGVAEGDILKSEVSEANADVQKE-----VISG	284
Qy	140	SLEKDLOEKY-GEPTQALLF--SAMRHVDHLDRINPQFQKSV-----	180
Db	285	QOEBQIEALIEGTAPVEVKETEVLTLEKDTLIDLEQPAQVPVAAEALPGVEAAEALV	344
Qy	181	-----ASDVFLAVESYRL-LAKOIDOPLH-----LGITAGARGSAVKSAIL-	224
Db	345	PSLEENKLGVEVVAPEAQOLIESAPVSPAPQPESTVLGVTE-----GDLKSEVSEADA	398
Qy	225	GLLISIEGIDTLRVSLAADPVEEIKVGF-----ILKSLIRSRGINFIACPTCSRQ	276
Db	399	GMOQEAQISD--QETQATEVEKVEVSEVETKEPEVEILLEGTLLIDLEQPAQVPVAAE	456
Qy	277	EFVDYIGVYVNL-----PQRLIEDITTPMDVSIIGCV--VNGGEALVSTLVGYNGKKSGL	329
Db	457	ELPGVEAAEALVPSLEENKLGVEVVAPEAQOLIESAPVSAVQPESTVLGVTEGDLKSEV	516

US-10-306-905-16

6.88; Score 126; DB 12; Length 533;

Best Local Simultaneously 23.4%; Field NO. 0.00066;
Matches 90; Conservative 57; Mismatches 115; Indels 122; Gaps 22;

```

QY 14 RIVYGNPVGICGAPLAVOSMNTTRTDEATVNOIKLERYGADIVRVSVPFMDAA--EA 71
Db 7 RILIFDTTLRDE-----QSPGATLT--VEEKLSIRALLARLVGDIIIEAGFPASGDEBA 60
QY 72 FKLIKOOV--NVBLV-----ADHIFDYRIALK 96
Db 61 VQKIAQVFTGNGEIVICGLARATOKDIIKAAEALPKPAKHRIHTFLATSDHLEHKTKT 120
QY 97 VAE-----YG---VDCLRINPGNIG--NEBRIMVYDCAKDKNIPRIIGNAGSL 141
Db 121 RAEVLAIVPEWVAAYAKSLVNDIEFSPEDAGRSDFEFLVQVLEAA-----ISAGAT 170
QY 142 EKDLQEKGEPTFPQ---ALLSAMSAMVHDLRLNDFQKVSYKA--SDVFLAVESYRLAK 197
Db 171 TINIPDVIYGTTPPAEYKALIKIGIADNP-----NIDQALISVGHNDUGLAVANTLEAVK 225
QY 198 QIDOPHLGITEAGARSAGVAKSAIGLILSEGIQDTRLVSLADPVEEIKVGEFDIJS 257
Db 226 -----NGARQ--LECTI-----NGIGE--RAGNAA--LEE-----LVWA 253
QY 258 LRIRSRGIN-FLACPTCSRQEPDVIQYVNA-LEQGLEDIITPMDSYIIGCVNGGEALV 315
Db 254 LHVRSRYFNPLGRPADSTAPLTINIDITHIATYSRLVSELGTGMWQ-----PNKAIV 305
QY 316 STLGVGNGKKSGLYEDGVRRDRL 339
Db 306 ---GANAFAHESGIHQDGVLNKKL 326

```

RESULT 7

```

US-10-306-905-14
/ Sequence 14, Application US/10306905
/ Publication No. US20030167513A1
/ GENERAL INFORMATION:
/ APPLICANT: MOURAD, GEORGE S.
/ APPLICANT: JUNK, DAMIAN J.
/ TITLE OF INVENTION: SELECTION AND USE OF ISOPROPYLMALATE SYNTHASE (IPMS)
/ TITLE OF INVENTION: MUTANTS DESSENSITIZED IN L-LEUCINE NEGATIVE FEEDBACK
/ TITLE OF INVENTION: CONTROL
/ FILE REFERENCE: 3220/93981
/ CURRENT APPLICATION NUMBER: US/10/306,905
/ CURRENT FILING DATE: 2002-11-27
/ PRIOR APPLICATION NUMBER: 60/339,895
/ PRIOR FILING DATE: 2001-11-30
/ NUMBER OF SEQ ID NOS: 45
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 14
/ LENGTH: 533
/ TYPE: PRT
/ ORGANISM: Microcystis aeruginosa
US-10-306-905-14

```

[illegible]

```

QY      144 LEEDUOEKGEPTPO---ALLSESAMRHVDHLDRNFQFKVSVKA-SDVFLAV-----ESY 192Z
Db      171 TVNVPDVTGYTTPBEPFALLRGIKENP-----NIDQAIISVGHDDLGIAVANFLAV 225S
QY      193 RLIAKOIDOPHLGITTEGKAGSAGVSAIGLILSBGIGDTLRVSLADPVEEIVGF 252Z
Db      226 KNGARPLECTIN-GIGERAGNAS-----LEE-----250C
QY      253 DILKSLRIRSGIN-FIACPTCSROEPFVIGTVNALE-ORLEDIITPMDVSIIGCVNCP 310I
Db      251 -LYMALHVRKRSIFNEFLGRPAESTPEPLTKITKEIYRSLVSNLTGMIVQ-----P 300I
QY      311 GEAIVSTLGVTGNGKSGLIEDGVKORL-----DNNIMIDOLEARIRAKAS 357Z
Db      302 NKAIIV---GANAFAHESGIHODGVKHLKLTYEINDAESIGLTNNQIVGLKLSGR-NAFRS 357Z
QY      358 OLDE 361
Db      358 RLDE 361 -

```

RESULT 8

```

US-10-156-761-10088
Sequence 10088, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156, 761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 10088
LENGTH: 1046
TYPE: prt
ORGANISM: Streptomyces avermitilis
US-10-156-761-10088

```

Query Match 5.8%; Score 108; DB 15; Length 1046;

Best Local Similarity 21.8%; Pred.NO. 0.13;
Matches 86; Conservative 48; Mismatches 139; Indels 122; Gaps 18.

QY	11	KSRRIVGVNVPJIGDGA-PLAVOSMTNRTTVEALVNOIKKLEBVGADIVRVSVPITDAA	69
Db	614	KSRTDIALIVVANDGMVPTQTEALNHAAADVPIV- AVNKIDVGDGAD-----PT-----	666
QY	70	EAFKLKQOVNPLVADIHEDYRIALKVAEYGVDCLRINPQNGNEERIKRVVDCARDKN	123
Db	663	---KVMGOLTEBGLVAE-----EYGGD-----TMEFV-----	686
QY	130	IPIRIVNNGSP-----EKDLOEKYGEPTQALLSESAMRHVDHLDRTNDFQKV-	178
Db	687	ISAKQGINESLLEAVVLTADASLDLRANPEPDAGIATIES-----HLDGRBAVATVL	740
QY	179	---SVKASDVFLAVESYRLAKOIQDPLHGIQTEAGARGAYKSAIGLGLLSEGI	234
Db	741	VQSGTGRVGDPMVVGDAYGRVAMDDGGE-NVEAGSGTEVLV-----LGLTVNPAGD	794
QY	235	TLAVSLAAPVPEIKVGPDIKLSLRIRRGINF-----ACPTCSQGEFDVI	281
Db	795	NFLV-----VVEDIRARQIAEKRRARRRNAPFARGVRFSLENUDEKLKAGLVQELNLI	848


```

APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIORITY APPLICATION NUMBER: JP 99/377484
PRIORITY FILING DATE: 1999-12-16
PRIORITY APPLICATION NUMBER: JP 00/159162
PRIORITY FILING DATE: 2000-04-07
PRIORITY APPLICATION NUMBER: JP 00/280988
PRIORITY FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 5705
LENGTH: 378
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5705

```

[illegible]

```

; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 256
; LENGTH: 387
; TYPE: PRN
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-256

```

	Query Match	40.3%; Score 751.5; DB 10; Length 387;
	Best Local Similarity	42.7%; Pred. No. 2,66-68;
	Matches	156; Conservative 64; Mismatches 134; Indels 11; Gaps 2
Qy	9 RRSKRTIYGVNPIGDAPIAVOSMTNRITTDVEATVNOIKALERVGADIVRVSPTMDA	68
Dd	18 RRATRQLMVGNGVGSDBPVSVOSGMTCTKTHDVA NSTLQQIAELMTAGCDIVRVACPRQED	77
Qy	69 AEAFLKIQQOVNVVLVDIHFDYIALKVAEYVDCIRINFGNIG-N-EERTRMVDDCARD	127
Dd	78 ADALAEARHSQIPFVVADIHFPRIYPAALIDAGCAAAVRVNPNIKIEFGRGVEVAKAGA	137
Qy	128 KNIFIRIGVNAAGSLPEKLOEKYEGRPTFOALLSESAMRHVDDHLDRLNPQFKYSVASDYFL	187
Dd	138 AGTIRIGVNAAGSIDKRFPMKEYGTATEALVESALMEASLFEEHGFGDIKTSVGHNDVPV	197
Qy	188 AVESYRLIAKOIDPPLHGITTEAGASGA VKSAIGHLLISEGICDTLRSLAADPYEE	247
Dd	198 MVAYELLARCDVPHLGVTENGPAPQGTTKSANVAGALLSRIGTIIRVLSAPREE	257
Qy	248 IKVGFEDILKSLIRISRGINFPLACFTCSRQEFDVIGTVNALEQRLEDITTPMDSIIIGCV	307
Dd	258 VKVGNQVLESINTLRPSLETIVSCSPCGRAQVDVYTLANEVTA GDGHDP LRVAVMGCV	317
Qy	308 NGPEBALVSTLGTGVGKKSGLYEDSVRKDRDLNNDMIIDOLEAIRAKASOLDERRIDV	367
Dd	318 NGPEAREADLVGASNGKGQOIFVRGEVIKTVPBAQIVETL-----IEEAMRLAA	367
Qy	368 QQVEK 372	
Dd	368 EMGEQ 372	
 RESULT 6 US-10-306-905-16		
	Sequence 16, Application US/10306905	
	Publication No. US2003016751A1	
	GENERAL INFORMATION:	
	APPLICANT: MOURAD, GEORGE S.	
	APPLICANT: JUNK, DAMIAN J.	
	TITLE OF INVENTION: SELECTION AND USE OF ISOPROPYLMALATE SYNTHASE (IPMS)	
	TITLE OF INVENTION: MUTANTS DESSENSITIZED IN L-LEUCINE NEGATIVE FEEDBACK	
	FILE REFERENCE: 3220/93381	
	CURRENT APPLICATION NUMBER: US/10/306,905	
	PRIOR FILING DATE: 2002-11-27	
	PRIOR APPLICATION NUMBER: 60/339,895	
	PRIOR FILING DATE: 2001-11-30	
	NUMBER OF SEQ ID NOS: 45	
	SOFTWARE: PatentIn Ver. 2.1	
	SEQ ID NO 16	
	LENGTH: 533	
	TYPE: PRT	
	ORGANISM: Synecchococcus PCC6803	

```

Oy 127 KNPIRIRGVNAGSLKENDOEKXGEPTPOALLIESARHVDHLDRLNEODQFKVSVKASDVF 186
Db 137 EHGTRIRGVNAGSLDRLLLEKXKATPEALVESALMEKSLFEEDHPRIKIKSVKANDVF 196
Oy 187 LAVESYRLIAKQIDOPHLIGITEAGARGAVKSAIGLILLSEIGDITLRYSLADPVE 246
Db 197 VMVNAVYRLIAQCDVPLHIGVTEAGPAFGQITKSAVAFALLSEIGDITRVSLSAPVE 256
Oy 247 EIKVGFEDILKSIRSRGINFIACPTCSQOEEDVIGTVALRQRREDITTPMDVSLIGCV 306
Db 257 EIKVGIQLLESINTLKRQGEIIVSCBSCGAQVDVYKLAEEVATGEGMVPPLRVAVMGCV 316
Oy 307 VNGPEALAVSTLGVGVGNKKSGLYEDGVAKRDLDDNNMDIDOLEARIRAKASQLEARRID 366
Db 317 VNGPEAREADGVASGNGKGQIFVKGEIKITVPESKIYETL-----IEEMKLI- 365
Oy 367 VQOVE 371
Db 366 AEQME 370

```

```

RESULT 2
US-09-881-752A-196
Sequence 196, Application US/09881752A
Patent No. US20020115078A1
GENERAL INFORMATION:
APPLICANT: Kleantous, Harold
APPLICANT: Al-Garawi, Amal
APPLICANT: Miller, Charles
APPLICANT: Tomb, Jean-Francois
TITLE OF INVENTION: Identification of Polynucleotides
TITLE OF INVENTION: Encoding No. US20020115078A1 Helicobacter Polypeptides in the
FILE REFERENCE: 061332/041002
CURRENT APPLICATION NUMBER: US/09/881,752A
CURRENT FILING DATE: 2001-06-15
PRIOR APPLICATION NUMBER: US 08/833,457
PRIOR FILING DATE: 1997-04-01
NUMBER OF SEQ ID NOS: 370
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 196
LENGTH: 359
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-881-752A-196

Query Match      42.6%; Score 794; DB 10; Length 359;
Best Local Similarity 45.8%; Pred. No. 1e-72;
Matches 162; Conservative 73; Mismatches 117; Indels 2; Gaps 2;

QY   RRSKRIIVGNPIEDGAPIVAVOSMTNTRTTTVEALVNOIKALERGADIVRVSPTMDA 68
DB   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   9 RRKTQKIFIGVAIGDAPISQTSMTFXTADIESTPKNDIDRLKLGAQLVRVAASNEKD 64
DB   S RVTKQKIFIGVAIGDAPISQTSMTFXTADIESTPKNDIDRLKLGAQLVRVAASNEKD 64
QY   AEAFLTIKQOVVVPIADIHFDYRIATLKVAAEYGVDCLRINPGNIINEERIRRWVDCARDK 128
DB   65 ALALKELKKVSPPLPADIHFFHYKRAL-TAASVDAIRINPNIKSKEIKAKVAVDDCKRK 123
QY   129 NIPRIGVAGSLEKLDOKYEKGEPTPOALLSESAMRHVDHLDRIINDFQFVSVKASDVFLA 188
DB   124 NIPRIGVAGSLEKLDOKYGF-PFKGVVESALYNAKLLIEDLFNFRTSLKASDVIRT 182
QY   189 VESYRLLAKQIDQPLHLGITTEAGGARSGAVKSAIGILLSSRGITTLRVLSAADVVEEI 248
DB   183 IEAYMPLRPVLVIYPHLGVTEAGNLFSSTSIKSAAMALGGLMBEGIGTMVTSITGELENBI 242
QY   249 KVGFIILSLRISRSGINFIACPTCSRQSFYIVGTVALAEQRLEDIITPMDSIIIGCVNN 308
DB   243 KVARRAILNRHSGLKKGINWITSCTCGRIIANLVDAIKYVKRSLSHKTPDLDISWGCVNN 302
QY   309 GGPEALVSTLGVTGNKKSGLYEDGVKRDRLDNNDIMIQLLEARIRAKASQLDEA 362

```

```

Db          303 ALGSAKHADMAIAFGNRSGLIIKEGKVIHKLAEXDLPETFVIEVNNLAKEREKS 356

RESULT 3
US-10-156-761-9186
; Sequence 9186, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9186
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9186

```

```

Query March 42.1% Score 786.5; DB 15; Length 385;
Best Local Similarity 44.4%; Pred. No. 6.6e-72;
Matches 164; Conservative 71; Mismatches 121; Indels 13; Gaps 4

Oy 6 PI-QRRKSTRIVGNVPIGDGAPIAVOSMTNRTTDDVATVNOIKALERGVADIVRSVP 64
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 15 PIARRRVSRCIOVGPLAVGTAFAVSVOSMTTTRTSDIGATLQOIATELTASGQIVRVACP 74
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 65 TMDAAEAFKLIKQOVNPLVVDIHFDVRLAKVAEYGDCLRINPNTGN--EERIIMVVD 123
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75 TODDADLAVIARSKQIPVADHIFQPKYVPAIEAGCAAVRVPNGNKQFDKKKEIAR 134
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 124 CARDNNPIRIGVAGSLKXDLOEKYGEPTPOALLSAMRHVDHLDRLNFDQFVSVKAS 183
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 135 AAKHGHPPIRIGVAGSLDRRLLEKYGKATPEALVESALTEASLFEEDHFDIKTSVGN 194
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 184 DVFLAVESYRLAAQIQPLHGLITTEAGARSGAVKSAISGLLLSEIGITLRYSLAAD 243
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 195 DPVVMVAVAYRGLAACQCYPLHLGVTGEPAPQGIKSAVARGALLSEIGITIRVSLAP 254
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 244 PVEEIKYGFPIDKSLRIRSRGINFIACPSCRSRQFVDVIGTVNALRLEQRLDITTPMDYSII 303
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 PVEEIKYGIQLBESLGRQRRLLEIVSCPSGCRADVVKLAEEVYTAGLEGNEVPLRVAVM 314
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 304 GCVTVNGPEALVSTLGTGNGKKSGLYEDGVKRLDNNDMITDLEAIRAKASQOLDAR 363
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 315 GCVVNGPEAREADLGVASGNGKQIFVKGIVIKTVPESKIVETL-----IEEAM 364
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 364 RIDVOQVER 372
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 365 KI-AEQWEDQ 372
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 4
US-09-738-626-5705
; Sequence 5705, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: WIZOGUCHI, HITROSHI
; APPLICANT: HAYASHI, MIKIRO

```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 17:16:44 ; Search time 48 Seconds

(without alignment)
1414,836 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866
Sequence: 1 MHNQAPIQRKRSTRIVGVN.....RAKASQDEARRIDVQVEK 372

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: Published Applications_AA:*
2: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	808	43.3	385	US-10-156-761-10098	Sequence 10098, A
2	794	42.1	359	US-09-881-752A-196	Sequence 196, App
3	786.5	42.1	385	US-10-156-761-9186	Sequence 9186, Ap
4	761	40.8	378	US-09-738-626-5705	Sequence 5705, Ap
5	751.5	40.3	387	US-09-712-363-256	Sequence 256, App
6	126	6.8	333	US-10-306-905-16	Sequence 16, App
7	110.5	5.9	533	US-10-306-905-14	Sequence 14, Appl
8	108	5.8	1046	US-10-156-761-10088	Sequence 10088, A
9	105.5	5.7	578	US-09-159-469-50	Sequence 50, Appl
10	105.5	5.7	578	US-09-798-042-50	Sequence 50, Appl
11	101	5.4	523	US-10-306-905-11	Sequence 11, Appl
12	99.5	5.3	573	US-10-306-905-15	Sequence 15, Appl
13	99	5.3	2368	US-09-815-242-5635	Sequence 5635, Ap
14	99	5.3	2368	US-09-815-242-12389	Sequence 12389, A
15	98.5	5.3	365	US-10-156-761-8947	Sequence 8947, Ap

16	96	5.1	504	9	US-09-815-242-10969	Sequence 10969, A
17	96	5.1	504	15	US-10-260-877-70	Sequence 70, Appl
18	94.5	5.1	290	9	US-09-815-242-5590	Sequence 5590, Ap
19	94.5	5.1	295	9	US-09-815-242-12428	Sequence 12428, A
20	94.5	5.1	483	12	US-10-301-997-51	Sequence 51, Appl
21	93	5.0	368	12	US-10-166-225A-62	Sequence 62, Appl
22	93	5.0	459	9	US-09-815-242-12703	Sequence 12703, A
23	93	5.0	1004	10	US-09-738-626-5676	Sequence 5676, Ap
24	93	5.0	2478	9	US-09-815-242-5816	Sequence 5816, Ap
25	93	5.0	2478	9	US-09-815-242-12967	Sequence 12967, A
26	92.5	5.0	307	15	US-10-128-714-4018	Sequence 4018, Ap
27	92	4.9	291	9	US-09-815-242-11266	Sequence 11266, A
28	92	4.9	498	10	US-09-738-626-5600	Sequence 5600, Ap
29	92	4.9	1798	10	US-09-845-583-8	Sequence 8, Appl
30	91.5	4.9	297	9	US-09-815-242-13258	Sequence 13258, A
31	91.5	4.9	332	9	US-09-815-242-13072	Sequence 13072, A
32	91.5	4.9	342	12	US-10-166-225A-63	Sequence 63, Appl
33	91.5	4.9	589	12	US-10-306-905-5	Sequence 5, Appl
34	91.5	4.9	756	10	US-09-738-626-6091	Sequence 6091, Ap
35	91	4.9	652	15	US-10-102-806-667	Sequence 667, App
36	91	4.9	855	15	US-10-156-761-14612	Sequence 14612, A
37	91	4.9	1798	10	US-09-938-275-9	Sequence 9, Appl
38	90.5	4.8	455	9	US-09-815-242-5420	Sequence 5420, Ap
39	90.5	4.8	455	9	US-09-815-242-12128	Sequence 12128, A
40	90.5	4.8	477	9	US-09-815-242-11826	Sequence 11826, A
41	90.5	4.8	502	15	US-10-156-761-12534	Sequence 12534, A
42	90.5	4.8	584	15	US-10-156-761-13475	Sequence 13475, A
43	90.5	4.8	842	9	US-09-815-242-11950	Sequence 11950, A
44	90	4.8	233	10	US-09-971-536-71	Sequence 71, Appl
45	90	4.8	358	9	US-09-784-508-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-10-156-761-10098 Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHITOKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10098
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10098
Query Match 43.3%; Score 808; DB 15; Length 385;
Best Local Similarity 44.9%; Pred. No. 4,1e-74;
Matches 164; Conservative 70; Mismatches 119; Indels 12; Gaps 3;

QY 8 ORKRSRIYGVNVPIDGAPIVAVQSWTNRTRTDVETATQVAKLEVGADIVSVPTMD 67
DB 17 ERKRSQIOVGIVAVGAPVAVQSWTNRTRTDVETATQVAKLEVGADIVSVPTMD 76
QY 68 AAEAFPLIKQOVVPLVADIHFDYRALKVAEYGVDCIRINFGNIGN-EERIRMVDCAR 126
DB 77 DADALAVIRAKSQIPVIAIDHIFQPKVVFALIEGCAAVVNGNITKQFPDKYKELAKAK 136

```
; Sequence 2, Application US/09612964
; Patent No. 6403342
; GENERAL INFORMATION:
; APPLICANT: Gusevskiy Mikhail Markovich
; APPLICANT: Lunts Maria Grigorievna
; APPLICANT: Kozlov Yuriy Ivanovich
; APPLICANT: Ivanovskaya Lirina Valerievna
; APPLICANT: Voroshilova Elvira Borisovna
; TITLE OF INVENTION: DNA CODING FOR MUTANT ISOPROPYLMALATE SYNTHASE
; TITLE OF INVENTION: L-LEUCINE-PRODUCING MICROORGANISM AND METHOD FOR PRODUCING
; TITLE OF INVENTION: L-LEUCINE
; FILE REFERENCE: 193845050
; CURRENT APPLICATION NUMBER: US/09/612,964
; CURRENT FILING DATE: 2000-07-10
; PRIOR APPLICATION NUMBER: RU 99114325
; PRIOR FILING DATE: 1999-07-09
; NUMBER OF SEQ ID NOS: 12
; SEQ ID NO 2
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Escherichia coli
; US-09-612-964-2
```

```
Query Match 5.4%; Score 101; DB 4; Length 523;
Best Local Similarity 20.3%; Pred. No. 0.046;
Matches 83; Conservative 62; Mismatches 142; Indels 122; Gaps 18;
```

```
QY 12 STRIYGVNVPICGAPFAIVOSMTNRTTDEATVNOIKALERVGADIVRSVPTMDAA-- 69
D 2 SQOVIIIFDTLRDGEQALQASLS-----VKEKIQIALALERMGVDEVFVSPSPDF 55
QY 70 EAFELIKQOVN---VPLVADIHFDYRIA---LKVAEYGVDCLEI-----NPGNIGNE- 115
D 56 ESVQTIARQVKNRSVCLACVEKIDVVAESLKVAE---AFRIHTFATSPMHIATKL 111
QY 116 -----ERIRNVVDCAR-----DKNIPRIGVNAAGSLEKDLQEK 148
D 112 RSTLDEVIERAIVWVKRANVTDDVEFSCBDAGRTPIADLARVVEAIIAGATTINIPDT 171
QY 149 YGEPTP---QALLESAMRHYDHLDFQPKSVKA-SVFLAVESYRLAKQIDQPLH 204
D 172 VGYTMPPEFAGIISGLYERVP-----NIDKAIISVHTHDDLGLAVGN-SLAA-----VH 219
QY 205 LGITEAGGARSQAVKSAIGLLSEGIQPTLRVSLADPVEIKVGFDTLKSLRIRSRG 264
D 220 AGARQVEGANV-----GIGERAGNCSLEEVIMAIKVRKDI----- 255
QY 265 INFIACTCSROEDVIGTVNALE-ORLEDIITPMVSIIGCVVNGPEALVSTLGVTTG 323
D 256 -----NVHTAINHQEIWRTSQVLSQ-----ICNMPIPANKKAIVSGAF 293
QY 324 NKTSGLYEDVGRKDLNDNDM-----IDOLEAIRAKASQOLDEARRID 366
D 294 AHSSGIHQDGLNKNRENYEIMTPESIGLNOIQIOLNLTSGRAAVKVRMD 342
```

```
Search completed: November 23, 2003, 17:17:39
Job time : 23 secs
```

```

QY      277 EDVIGTVAL-----EQLIEDITPMDSIIICV--VNGEGALVSTLTGNGKKSGL 329
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      457 ELPGVEAAEAIVPSLEENKIQEVVAPEAQQLSAPFVSAPVOESTVLGTBGLNSEV 516
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY      330 -----YEDGVKRDLNDNMDIMDILEARIRAKASQLDEARRIDV 367
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      517 SVADAGMQGEAGISDQETQATEVEKEVSVESADGMQGL--VDV 561
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 13
US-09-106-582-50
; Sequence 50, Application US/09106582
; Patent No. 6306402
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/106,582
; FILING DATE: 28-JUN-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.439C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-4900
; TELEFAX: 206-682-6031
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 578 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-106-582-50

Query Match           5.7%; Score 105.5; DB 4; Length 578;
Best Local Similarity 19.5%; Pred. No. 0.018; Indels 123; Gaps 21;
Matches 91; Conservative 77; Mismatches 176;

QY      7 IQRRSTRIVGNVP-----IGDGA-----PIAVQSMTNTRT-TDV EATVNQ- 47
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     112 VQEEBGTYGYLINAPEKAVVRFFIKSAAEHPQTVPDSVESATSGVDTQEEGEIDQE 171
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY      48 IKALERVGAD-----IVRSVPTMDAAEF--KLTIQQVNV 82
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     172 APAIEVEETEBOEVLIEBGLIDLBPVAVPVVAEALPVGVEAMEAIVPSLEENKLOEV 231
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY      83 LVADIHPDRILKV---AEYGVDCLRINPNTIGNERIRMVDCARDKNIPRI GNAG 139
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     232 VVAPAQQLESAPESAPAPQESTVLGVAEGDKSEYSVANADVAKQK-----VISG 284
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY     140 SLEKDIOEKY-GEPTQALL--SAMRHVDHLDRINFDDQKFSV----- 180
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db     285 QOQGIAIALBEGTEAPEVKEETFEVLTKEDTLIDLBPVAVQVPVVAEALPGVEAAEAIV 344
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY     181 -----KASDVFLAESYRL-LAKQIDQPLH-----LGITEAGARGAGVKAISGL---- 224
          |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

```

Db      345 PSLEENKLGQEVVVAPEAQQLSEAPRQPESTYLGATYE-----GDKSEVSEVADA 398
Oy      225 GLLLSGIGDTLRVSLAADPVEEIKVGFED-----ILKSLIRSRGINFIACPTCSRQ 276
Db      399 GMOOEAGISD--QETQATEEVEKVEVSERTKTEPEVILEEGTILIDLECPVAPVVAEA 456
Oy      277 EFDVIGVYVNL-----EQRLIEDITPMDVSIIGCV--VNGPGALVSTIGVTKGNKSKGL 329
Db      457 ELDPGEVAEAEIVPSLEENKLGQEVVVAPEAQQLSEAPRQPESTYLGATYE-----GDKSEV 516
Oy      330 -----YEDGVKRDRLDNNMDIDLEARIKAKASQLEARRIDV 367
Db      517 SVENDAGMOEAGISDQETQATEEVEKVEVSVEADAGMOEQL--VDV 561

RESULT 14
; Sequence 29387, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ. ID NOS: 33142
; SEQ ID NO 29387
; LENGTH: 465
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29387

Query Match      5.4%; Score 101, DB 4; Length 465;
Best Local Similarity 21.8%; Pred. No. 0.038;
Matches 81; Conservative 43; Mismatches 122; Indels 126; Gaps 15;

Oy      25 GAPIAV-QSMTNRTTDDVEATVNOIKALERVGADIVRSVPTMDABAPFLIKQOVNPL 83
Db      59 GLEIALAETTLVALADLDLE---EDRDNRKVGEDL-----QGDVAVARGAVDQOVPAQ 108
Oy      84 VADIHFDYRIALKAVEYGVDCLRINPGNIGNEERIRIMVDC-----125
Db      109 LLE-----RLAMP-GDQGVDLVV--GVRRHHLHVVVNGLAHGSDIVVGGQDVLDAFA 160
Oy      126 -----RDKNIPTIR---IGVNAQSLEKIDQOEKYGGEPTPALLBSA- 161
Db      161 MVEADEFLDRLGIVGFVFDRLDADLPARCGHGAEOAGEILAFDVEVNLAEVGDALVETGP 220
Oy      162 -----MRHVDHLDRL-----NPGQFVSKASDVFLEAVESYRLLAQIQPRL 204
Db      221 DVHLAALDVGVVVDVQADRVVVVGAAPFEELEVDVTDAAVAAVDEIQRTDAFQACD 280
Oy      205 LGITEAGGA--RSQAVVSAIGLGL--LLSEGIQDTLRVSLAADPVEEIKVGFEDILKSLR 259
Db      281 VQLAEVGVAAHQIGTLGFGVGGGLSVLHPEGHGAGARVLLBEE-----324
Oy      260 IRSRGINFIACTPSCROEFPVITGVNALBEQRLIEDITPMDVSIIGCVVNGPGEA----- 313
Db      325 -----FVDMPPRAAVEHDV-----DIVLLEQEDFLGTMLGGFGEAHVHEQL 365
Oy      314 --LVSTIGVTGG 323
Db      366 AOLLDALGVGGG 377

```

OY	217	WKSJNIGGLISEGDLTFLVSLADPVEELKVGPDIIKSLIRSRGNGINFCPTCSRO	276
Db	645	QVLLAVGVG-IETGVAENVLIIVVRLGPEQRAVENQAI-----VEALIAAQVAGQ	693
OY	277	EFDVI-----GTVALLEQRLBEDIITPMDSIIGCVVNGEGALVSTLG-VTGGNKK	326
Db	694	VAAVASWMLIARSTGALGIDPRRIE-AAHPEAVPAVAGAADPGALQGMGAVTGEER	752
OY	327	SGLY-----EDGVRKDR-----LDNNMDIQLFARI--RAKASQLDERRIV	367
Db	753	LGALAAVAFGEDLDHPADGLRAVQAGTRPADHLDALDLHROVLEGEAS---AGRADL	808
OY	368	QOVER 372	
Db	809	DAVDQ 813	
RESULT 11			
US-08-975-762-50			
Sequence 50, Application US/08975762			
Patent No. 6207169			
GENERAL INFORMATION:			
APPLICANT: Reed, Steven G.			
APPLICANT: Lodes, Michael J.			
APPLICANT: Houghton, Raymond			
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS AND			
NUMBER OF SEQUENCES: 73			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: SEED and BERRY LLP			
STREET: 6300 Columbia Center, 701 Fifth Avenue			
CITY: Seattle			
STATE: Washington			
COUNTRY: USA			
ZIP: 98104			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/975,762			
FILING DATE: 21-MAR-1997			
CLASSIFICATION: 424			
ATTORNEY/AGENT INFORMATION:			
NAME: Makl, David J.			
REGISTRATION NUMBER: 31,392			
REFERENCE/DOCKET NUMBER: 210121.439			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: 206-682-6031			
TELEFAX: 206-682-4900			
INFORMATION FOR SEQ ID NO: 50:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 578 amino acids			
TYPE: amino acid			
STRANDEDNESS:			
TOPOLOGY: linear			
MOLECULE TYPE: protein			
US-08-975-762-50			
OY	Query Match	5.7%; Score 105.5; DB 3; Length 578;	
Db	Best Local Similarity	19.5%; Pred. No. 0.018;	
	Matches 91; Conservative	77; Mismatches 176; Indels 123; Gaps 21;	
OY	7	TORRSTRIYGVNVP-----ICDGA-----PIAVGSMNTTRT-TDVEALVNO-	47
Db	112	VOEEGTGYTLINAPKAVVRPFKLEKSAEEPQTVDPBSVESATGSGVDTOEEQIDQE	171
OY	48	IKALERGVAD	
Db	172	APAIKEVEEHEBEVILIEGTLIDLEQPAVQVPAVMAELPGVBAALAIYPSLEENLQJEV	231
OY	83	LVADITHFDYRIALKY--AEYGVDCIRINPGINIGNEERIRMVAVDCCARDKNIPIRIGVNAQ	139

```

Db      232 VVAPEAQLQESAPAEVSAPAPQESTVLGVAEBDLKSEVSVEANADVAQNE-----VLSG 284
Qy      140 SLEKDLOEKY-GEPTPOLLE-SAMRHVDHLDRINFQFVSV----- 180
Db      285 QOQOEIALEAEGTAPAVEKETEVLKEDTLIDLEGQVAVPVAAEALPGVEAAEAIV 344
Qy      181 -----KASDVFLAVESYRL-LAKQIDQPLH-----LGITAGGARSAAVSAIGL----- 224
Db      345 PSLEENKLOEVVVAPEAQOLESAPAEVSAPAPQESTVLGVTE-----GDLKSEVSVEADA 398
Qy      225 GLTISEGIDTLRLSLADPVEELKVGPD-----ILKSLRIRSRGINFIACPTCSRQ 276
Db      457 ELPGVEAAEAIVPSELENKLOEVVVAPEAQOLESAPAEVSAPVQESTVLGVTEBGDLKSEV 516
Qy      330 -----YEDGVKDRDLNDNDIMIDOLEAIRAKASQDLBARRDV 367
Db      517 SVEADAGMOQEAQISDOETQATEBEVEKEVSVEADAGMOQEL--VDV 561

RESULT 12
US-09-295-028-50
/ Sequence 50, Application US/09295028
/ Patent No. 6277381
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Lodes, Michael J.
/ APPLICANT: Houghton, Raymond L.
/ APPLICANT: McNeill, Patricia D.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
/ TITLE OF INVENTION: AND TREATMENT OF EHRlichia INFECTION
/ FILE REFERENCE: 210121.439C4
/ CURRENT APPLICATION NUMBER: US/09/295.028
/ CURRENT FILING DATE: 1999-04-20
/ NUMBER OF SEQ ID NOS: 85
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 50
/ LENGTH: 578
/ TYPE: FRT
/ ORGANISM: Ehrlichia sp.
US-09-295-028-50

Query Match      5.7%; Score 105.5; DB 3; Length 578;
Best Local Similarity 19.5%; Pred.No. 0.018;
Matches 91; Conservative 77; Mismatches 176; Indels 123; Gaps 21

Qy      7 IORRKSRIYGVNP-----IGDGA-----PIAVQSMTNTRT-TDVEATVQ- 47
Db      112 VQEEEGGMLINPEKAVVAFRIEKSAAEPTQVDSVESATGSGVDVQOEBEIDOE 171
Qy      48 IKALERVGAD-----IVRSVPTMDAAEAF-KLIKQVNP 82
Db      172 APAIEEVEETEOEVLIEGTLIDLEQVAVQPVAAEALPGVEAAEAIVPSELENKLOEV 231
Qy      83 LVADIHHDYRLALV--AEYGVDCILINPQINQNEERIRNVVDCARQKXNPIRIQVNAQ 139
Db      232 VVAPEAQOLESAPAEVSAPAPQESTVLGVAEBDLKSEVSVEANADVAQNE-----VLSG 284
Qy      140 SLEKDLOEKY-GEPTPOLLE-SAMEHVDHLDRINFQFVSV----- 180
Db      285 QOQOEIALEAEGTAPAVEKETEVLKEDTLIDLEGQVAVPVAAEALPGVEAAEAIV 344
Qy      181 -----KASDVFLAVESYRL-LAKQIDQPLH-----LGITAGGARSAAVSAIGL----- 224
Db      345 PSLEENKLOEVVVAPEAQOLESAPAEVSAPAPQESTVLGVTE-----GDLKSEVSVEADA 398
Qy      225 GLTISEGIDTLRLSLADPVEELKVGPD-----ILKSLRIRSRGINFIACPTCSRQ 276
Db      457 ELPGVEAAEAIVPSELENKLOEVVVAPEAQOLESAPAEVSAPVQESTVLGVTEBGDLKSEV 516
Qy      330 -----YEDGVKDRDLNDNDIMIDOLEAIRAKASQDLBARRDV 367
Db      517 SVEADAGMOQEAQISDOETQATEBEVEKEVSVEADAGMOQEL--VDV 561

```

```

QY 134 IGVN-----AGLEKDJCKYGEFFFOALLLESAMRHVDHLRNFQFKVSVASVPLA 188
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db 414 --INHALLSGTTRIRPLSDAEQFPGRSLVE-----ILKRTTFVFL 453
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
QY 189 VESTYLLAKQIDQPLHLGITTEAGARGAIVKAGLGLLSEIGD----- 234
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db 454 DEIDGLKSAIDS---SVTE---LRLGVNGDNPILRLFNHVHGKSHIPGIDRSSLYEV 506
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
QY 235 -----TLRVSLAD-----PVEIKVGFPLDKSLRISRGINFACPSROEF 278
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db 507 NWALMYCTLSYQLNLQOERYFEMPIKA-TWMSDADDTFTIOALCI-----TRETLEY 559
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
QY 279 DVIGTVALBQRLIEDITPMDSYITGCYVN---GPGEALVSTLGVYGNKKGSLYEDGVR 335
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db 560 LFSNDIGOLPHHLQ-----SLERNILNMRNSNDGHRLETLATLGLNLIDLSEGRLL 611
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
QY 336 KDRLDNNMDIQLEARIKAKAS 357
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:
Db 612 PERLKEHD-----RIKLKAS 626
      .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:

```

RESULT 9

```

US-09-066-046-6
/ Sequence 6, Application US/09066046A
/ Patent No. 6204252
/
GENERAL INFORMATION:
/
APPLICANT: MURPHY, Cheryl
/ STOREY, James
/ BELTZ, Gerald A.
/ COUGHLIN, Richard T.
/
TITLE OF INVENTION: CHARACTERIZATION OF GRANULOCYTTIC
/ EHRLICHIA AND METHODS OF USE
/
NUMBER OF SEQUENCES: 65
/
CORRESPONDENCE ADDRESS:
/ ADDRESSEE: HALE AND DORR LLP
/ STREET: 60 State Street
/ City: Boston
/ STATE: MA
/ COUNTRY: United States
/ ZIP: 02109
/
COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Releasee #1.0, Version #1.30
/
CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/066,046A
/ FILING DATE: 24-Apr-1998
/ CLASSIFICATION: <Unknown>
/
ATTORNEY/AGENT INFORMATION:
/ NAME: Superko, Colleen
/ REGISTRATION NUMBER: 39,850
/ REFERENCE/DOCKET NUMBER: 106,941.155
/
TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617) 526-6000
/ TELEFAX: (617) 526-5000
/
INFORMATION FOR SEQ ID NO: 6:
/
SEQUENCE CHARACTERISTICS:
/ LENGTH: 578 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
/
US-09-066-046-6

```

Query Match	6.0%;	Score 111.5;	DB 3;	Length 578;
Best Local Similarity	18.6%;	Pred. No. 0.0039;		
Matches	88;	Conservative	77;	Mismatches 172;
				Indels 137;
				Gaps 20;
7 IQRRKSTRIVGNP-----IGDA-----PIAVQSMNTPT-TDVEATVQ--	47			

```

Db      112  V0EEBCTGTYLINADEKAVVREFKIEKSAEERPQTVDPEVSVAESATGSGVDT0EOE0IDOE 1711
Qy      48   IKALERVGADIVRVSVPTMDAAEAFKLIKQVNVVLVADIHEDYRIALTKVAEYGVCLRI 1070
Db      172  APAIEVEVTEBEQV---ILBEBTLLIDLEBPVAVQVVAEALP---GVNAEALIVESL-- 2233
Qy      108  NPGNIGNEERIRVVV-----DCARDKNPIR-----IGNVAGSLEKLOEKYEPTP 154
Db      224  -----BENKLOEVVAPEAQOLJESAPAEVSAPAPCESTVLGVAEGDLKSEVSVEANADVP 2777
Qy      155  QALJESARHVDHIDRLNFDQEKSV----- 180
Db      278  QKEVISGQEOEBIAEALBETPAPEVKEETEVLKEDTLIDLEQPVAVPVAEALPGV 337
Qy      181  -----KASDVFLAVESYRL-LAKOIDDPLH-----LGITBAGARGSAVKSA 221
Db      338  EAAEALIVPSLEENKLOEVVAPEAQOLJESAPAEVSAPAPCESTVLGYTE-----GDLKEE 391
Qy      222  IGL-----GLLISEGIGDTLRLVSLADPVPEELIKGVD-----ILKSLIRSRGINFIA 269
Db      392  VSVEADAGMQ0EAGISD--GEIQAITEEBEKVEVSYTEETBEPEVILBEBTLLIDLEBPVAVQ 449
Qy      270  CPTSGR0EFDVIYGVNAL-----B0REDDITTPMDVSIIGCV--VNGPEEALVSTLGYTG 322
Db      450  VPVVAEALPGEVAAEALIVPSLEENKLOEVVAPEAQOLJESAPAEVSAPAPCESTVLGYTE 509
Qy      323  GNKKSGL-----YEDGVKORLNDNMDIDLEARIIRAKASOLDREARRIDV 367
Db      510  GDLKSEVSVEADAGMQ0EAGISDQETQAITEEBEKVEVSVEADAGMQ0EL--ADV 561

```

RESULT 10

```

US-09-252-991A-30838
Sequence 30838, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERGININOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252.991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074.788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094.190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ. ID NOS: 33142
SEQ. ID NO 30838
LENGTH: 960
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30838

```

Query Match	5.7%	Score 106.5	DB 4	Length 980
Best Local Similarity	20.7%	Pred. 0.033		
Matches	88	Conservative	65	Mismatches 159
			Indels	113
			Gaps	20

QY	16	YGVNVP	IGDGA	PAV	QSM	TN	TR	TT	DT	VT	AT	Y	NQ	I	A	L	E	R	V	G	A	D	I	V	R	V	S	V	P	T	M	D	A	E	F	K	L	I	75														
Db	434	FV	D	V	D	V	R	P	P	G	A	P	R	S	R	T	R	I	V	A	D	Q	V	D	V	D	V	G	T	---	E	G	I	R	P	V	P	A	R	T	A	P	D	I	Q	A	A	S	L	E	G	-	487
QY	76	KQ	O	N	N	V	P	L	V	A	D	I	H	---	D	Y	R	---	A	L	K	V	E	Y	G	V	D	C	I	R	M	P	G	N	I	G	N	E	E	R	I	R	M	V	D	C	124						
Db	488	--	Q	V	G	I	A	N	V	G	V	E	F	E	G	R	L	E	C	R	A	V	D	G	E	A	R	V	A	---	R	V	A	P	G	A	A	G	V	H	A	E	L	L	V	V	539						
QY	125	A	R	D	K	N	I	P	I	R	I	G	N	A	G	S	L	E	K	D	L	O	E	K	Y	G	E	P	T	---	P	O	A	L	L	E	S	---	A	M	R	H	V	D	H	L	R	N	F	D	174		
Db	540	T	A	Q	A	Q	L	P	T	---	L	A	T	V	A	E	L	V	G	E	P	A	F	Q	P	R	V	D	A	B	G	F	A	A	D	L	V	L	P	V	A	E	G	Q	M	A	F	G	591				
QY	175	Q	F	K	V	S	V	A	S	V	F	L	A	V	E	S	T	R	L	L	A	K	O	I	D	P	L	H	G	T	---	T	E	A	C	A	R	S	216														
Db	592	Q	G	G	V	V	I	P	G	O	A	V	A	T	G	I	E	A	---	R	V	E	T	A	I	I	G	I	O	L	V	L	A	P	L	E	G	H	O														

QY 241 AADPVEIKVFDILKSLRISRGINFACPTCSROEFDVIGTVALNLEQLEDDITPMV 300
 DB 419 AADPVEIKVFDILKSLRISRGINFACPTCSROEFDVIGTVALNLEQLEDDITPMV 478
 QY 301 SIIGCVNNGEALVSTLGTGNGKSGLYEDGVKRDLDNNMIDOLEARIRAKASOLD 360
 DB 479 AVIGCVNNGEALVSTLGTGNGKSGLYEDGVKRDLDNNMIDOLEARIRAKASOLD 537
 QY 361 EA 362
 DB 538 EA 539

RESULT 6

US-09-328-352-7906
 ; Sequence 7906, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 7906
 ; LENGTH: 378
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii
 ; US-09-328-352-7906

Query Match 64.0%; Score 1195; DB 4; Length 378;
 Best Local Similarity 63.8%; Pred. No. 6.7e-122; Indels 0; Gaps 0;

Matches 231; Conservative 65; Mismatches 66; Indels 0; Gaps 0;

QY 6 PLORRKSTRIVGNVPIGDPAPIAVOSMTNRTTVEATVNOIKALERVAGDIVSVPT 65
 DB 12 PLORRKSTRIVGNVPIGDPAPIAVOSMTNRTTVEATVNOIKALERVAGDIVSVPT 71
 QY 66 MDAAEAFKLIQOONVPLVADIDHFDYRIALKVAEYGVDCIRINPNGINEERIRMVVDC 125
 DB 72 MDAAEAFKLIQOONVPLVADIDHFDYRIALKVAEYGVDCIRINPNGINEERIRMVVDC 131
 QY 126 RDKNIPRIKIVNGSLERKDOEKYGEPTPALLSARHVDHLDRLNFDPFKVSASDV 185
 DB 132 RDKNIPRIKIVNGSLERKDOEKYGEPTPALLSARHVDHLDRLNFDPFKVSASDV 191
 QY 186 FLAVESYRLAKOIDPLHIGTEAGARSGAVSAIGLILSEIGDILRSLADPV 245
 DB 192 FLAVESYRLAKOIDPLHIGTEAGARSGAVSAIGLILSEIGDILRSLADPV 251
 QY 246 EEIKVGFILKSLRISRGINFACPTCSROEFDVIGTVALNLEQLEDDITPMVDSIIGC 305
 DB 252 EEIKVGFILKSLRISRGINFACPTCSROEFDVIGTVALNLEQLEDDITPMVDSIIGC 311
 QY 306 VVNGPGALVSTLGTGNGKSGLYEDGVKRDLDNNMIDOLEARIRAKASOLD 365
 DB 312 VVNGPGALVSTLGTGNGKSGLYEDGVKRDLDNNMIDOLEARIRAKASOLD 371
 QY 366 DV 367
 DB 372 EI 373

RESULT 7

US-09-198-452A-389
 ; Sequence 389, Application US/09198452A
 ; Patent No. 6559294
 ; GENERAL INFORMATION:
 ; APPLICANT: Griffiths, R.
 ; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
 ; thereof and uses thereof, in particular for the diagnosis, prevention
 ; TITLE OF INVENTION: and treatment of infection

FILE REFERENCE: 9710-003-999
 ; CURRENT APPLICATION NUMBER: US/09/198,452A
 ; CURRENT FILING DATE: 1998-11-24
 ; NUMBER OF SEQ ID NOS: 6849
 ; SEQ ID NO 389
 ; LENGTH: 621
 ; TYPE: PRT
 ; ORGANISM: Chlamydia pneumoniae
 ; US-09-198-452A-389

Query Match 25.2%; Score 470.5; DB 4; Length 621;
 Best Local Similarity 38.9%; Pred. No. 2.5e-42;
 Matches 111; Conservative 55; Mismatches 84; Indels 35; Gaps 5;

QY 9 RKXSTRIVGNVPIGDPAPIAVOSMTNRTTVEATVNOIKALERVAGDIVSVPTMA 68
 DB 20 RKXSTRIVGNVPIGDPAPIAVOSMTNRTTVEATVNOIKALERVAGDIVSVPTMA 79
 QY 69 AEAFLIKQO-----VVNPLVADIDHFDYRIALKVAEYGVDCIRINP----- 111
 DB 80 AEAFLIKQO-----VVNPLVADIDHFDYRIALKVAEYGVDCIRINP----- 138
 QY 112 -----IGNEERIRMVVDCARDKXNIPRIGVNGSLERKDOEKYGEPTPALLS 160
 DB 139 IYTESVAGSLRLERKFAVLEKCKRIGKAMRIGVNGSLERKDOEKYGEPTPALLS 197
 QY 161 AMRHVDHLDRLNFDPFKVSASDVFLAVESYRLAKOID-----OPLHIGTEAGAR 215
 DB 198 AMRHVDHLDRLNFDPFKVSASDVFLAVESYRLAKOID-----OPLHIGTEAGAR 257
 QY 216 GAVKSAIGLILSEIGDILRSLADPV 260
 DB 258 GAVKSAIGLILSEIGDILRSLADPV 302

RESULT 8

US-09-252-991A-29105
 ; Sequence 29105, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252,991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074,788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094,190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 29105
 ; LENGTH: 1058
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-29105

Query Match 6.0%; Score 112.5; DB 4; Length 1058;
 Best Local Similarity 21.5%; Pred. No. 0.0084;
 Matches 95; Conservative 50; Mismatches 140; Indels 157; Gaps 21;

QY 6 PLORRKSTRIVGNVPIGDPAPIAVOSMTNRTTVEATVNOIKALERVAGDIVSVPT 65
 DB 252 PLORRKSTRIVGNVPIGDPAPIAVOSMTNRTTVEATVNOIKALERVAGDIVSVPT 296
 QY 66 M-DAEAFKLIQO-----VVNPL-VADIDH-----FDYRIALKVAE 99
 DB 297 M-DAEAFKLIQO-----VVNPL-VADIDH-----FDYRIALKVAE 354
 QY 100 YGVDCIRINP-----IGNEERIRMVVDC-----ARDKIPR 133
 DB 355 YGVDCIRINP-----IGNEERIRMVVDC-----ARDKIPR 413

Db 1 MHESP1KRKRSRTIYGVNVPIDGAP1AVQSMNTTRTTTVEATVRQIQSLERVGDIYR 60
Qy 61 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRM 120
Db 61 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRQ 120
Qy 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPQALLSARHVDHDLRLNFDQKVSU 180
Db 121 VVDSARHNIPIRIGVAGSLKEDLOEKYGEPTPEALVESAMRHVDLRLNFDQKVSU 180
Qy 181 KASDVFLAVGSYRLAKQIDQPHLGITTEAGARSAGVSAIGLGLLSGIGDTLRVSL 240
Db 181 KASDVFLAVGSYRLAKQIDQPHLGITTEAGARSAGVSAIGLGLLSGIGDTLRISL 240
Qy 241 AADPVEIKVGFPIILKSLRIRSRGINFIACPTCSROEFVIGTVNALQORLEDIITPMV 300
Db 241 AADPVEIKVGFPIILKSLRIRSRGINFIACPTCSROEFVIGTVNALQORLEDIITPMV 300
Qy 301 SIIGCVNGBGEALVSTLGVTKGSKSLYEDGVR-KDRLNDMDIDLEARIKASQSL 359
Db 301 SIIGCVNGBGEALVSTLGVTKGSKSLYEDGVRKDERFDNDNIIDLEAKIRAKAAML 360
Qy 360 DE 361
Db 361 DE 362

RESULT 4
US-09-170-187-5
Sequence 5, Application US/09170187
Patent No. 6383745

GENERAL INFORMATION:

APPLICANT: Rafter, Philip N.

TITLE OF INVENTION: Methods For Screening For Anticicrobials
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medlen & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94104

COMPUTER READABLE FORM:

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/170,187

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/827,190

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Carroll, Peter G.

REGISTRATION NUMBER: 32,837

REFERENCE/DOCKET NUMBER: CASE-02443

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 705-8410

TELEFAX: (415) 397-8338

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 365 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-170-187-5

Query Match 87.0%; Score 1622.5; DB 4; Length 365;
Best Local Similarity 88.4%; Pred. No. 1.2e-168;

Matches 320; Conservative 24; Mismatches 17; Indels 1; Gaps 1;
Qy 1 MHNPAP1ORRKSRTIYGVNVPIDGAP1AVQSMNTTRTTTVEATVRQIKALERVGADIYR 60
Db 1 MHESP1KRKRSRTIYGVNVPIDGAP1AVQSMNTTRTTTVEATVRQIQSLERVGDIYR 60
Qy 61 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRM 120
Db 61 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRQ 120
Qy 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPQALLSARHVDHDLRLNFDQKVSU 180
Db 121 VVDSARHNIPIRIGVAGSLKEDLOEKYGEPTPEALVESAMRHVDLRLNFDQKVSU 180
Qy 181 KASDVFLAVGSYRLAKQIDQPHLGITTEAGARSAGVSAIGLGLLSGIGDTLRVSL 240
Db 181 KASDVFLAVGSYRLAKQIDQPHLGITTEAGARSAGVSAIGLGLLSGIGDTLRISL 240
Qy 241 AADPVEIKVGFPIILKSLRIRSRGINFIACPTCSROEFVIGTVNALQORLEDIITPMV 300
Db 241 AADPVEIKVGFPIILKSLRIRSRGINFIACPTCSROEFVIGTVNALQORLEDIITPMV 300
Qy 301 SIIGCVNGBGEALVSTLGVTKGSKSLYEDGVR-KDRLNDMDIDLEARIKASQSL 359
Db 301 SIIGCVNGBGEALVSTLGVTKGSKSLYEDGVRKDERFDNDNIIDLEAKIRAKAAML 360
Qy 360 DE 361
Db 361 DE 362

RESULT 5
US-09-252-991A-22991

Sequence 22991, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfeld et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 22991

LENGTH: 547

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-22991

Query Match 73.6%; Score 1372.5; DB 4; Length 547;
Best Local Similarity 73.8%; Pred. No. 4.9e-141;

Matches 267; Conservative 46; Mismatches 48; Indels 1; Gaps 1;

Qy 1 MHNPAP1ORRKSRTIYGVNVPIDGAP1AVQSMNTTRTTTVEATVRQIKALERVGADIYR 60
Db 179 IHSASPIIRKRSKRIKIVGVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRM 238
Qy 61 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRM 120
Db 239 VSVPTMDAAEAFKLIKQOVNVPVAD1HFDYRIALKVAEYGVDCLRINPNCIGNERIRV 298
Qy 121 VVDCARDKNIPIRIGVAGSLKEDLOEKYGEPTPQALLSARHVDHDLRLNFDQKVSU 180
Db 299 VVDSARHNIPIRIGVAGSLKEDLOEKYGEPTPEALVESAMRHVDHDLRLNFDQKVSU 358
Qy 181 KASDVFLAVGSYRLAKQIDQPHLGITTEAGARSAGVSAIGLGLLSGIGDTLRVSL 240
Db 359 KASDVFLAVGSYRLAKQIDQPHLGITTEAGARSAGVSAIGLGLLSGIGDTLRISL 418

Db 61 VRVSFTMDAAEFKLIKQVNVPLVADHFDRIALKAAYGVDCFTLRINPQNGNEE 120
QY 117 RIRMVDCARDKNIPIRIGVNGSLEKDLQEKYGEPTPALLESAR--HVHDLRLNFD 174
Db 121 RIRMVDCARDKNIPIRIGVNGSLEKDLQEKYGEPTPALLESARFTHVHDLRLNFD 180
QY 175 QFVSVKASDVFLAVESYLLAKQIDQPHLGTTEAGARSAGVKSAT--GIGLLISEGI 232
Db 181 QFVSVKASDVFLAVESYLLAKQIDQPHLGTTEAGARSAGVKSATFTGLLISEGI 240
QY 233 GDTLRVSLAADPVEEIKVGFIDILKSLRIRSRGINFIACPTCSROEPDVI--GTVNALEOR 290
Db 241 GDTLRVSLAADPVEEIKVGFIDILKSLRIRSRGINFIACPTCSROEPDVI--GTVNALEOR 300
QY 291 LEDITPMVDSIIGCVNPGGALVSTLGVGKNGKSGLYEDGVKRDRLD--NNDMIDOL 348
Db 301 LEDITPMVDSIIGCVNPGGALVSTLGVGKNGKSGLYEDGVKRDRLDFTNNDMIDOL 360
QY 349 EARIRAKASQLEARRIDVQOVER 372
Db 361 EARIRAKASQLEARRIDVQOVER 384

RESULT 2

US-09-170-187-6
Sequence 6, Application US/09170187
Patent No. 6383745
GENERAL INFORMATION:
APPLICANT: Rather, Philip N.
TITLE OF INVENTION: Methods For Screening For Antimicrobials
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/170,187
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/827,190
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 384 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-170-187-6

Query Match 96.5%; Score 1800; DB 4; Length 384;
Best Local Similarity 96.9%; Pred. No. 4; 9e-188;
Matches 372; Conservative 0; Mismatches 0; Indels 12; Gaps 6;
QY 1 MHNQAPIQRKSTRIVGVNVPIDGAPIAVQSWTNTTRTDTVEATV--NQIKALERVGADI 58

Db 1 MHNQAPIQRKSTRIVGVNVPIDGAPIAVQSWTNTTRTDTVEATV--NQIKALERVGADI 60
QY 59 VRVSFTMDAAEFKLIKQVNVPLVADHFDRIALKAAYGVDC--LRINPQNGNEE 116
Db 61 VRVSFTMDAAEFKLIKQVNVPLVADHFDRIALKAAYGVDCFTLRINPQNGNEE 120
QY 117 RIRMVDCARDKNIPIRIGVNGSLEKDLQEKYGEPTPALLESAR--HVHDLRLNFD 174
Db 121 RIRMVDCARDKNIPIRIGVNGSLEKDLQEKYGEPTPALLESARFTHVHDLRLNFD 180
QY 175 QFVSVKASDVFLAVESYLLAKQIDQPHLGTTEAGARSAGVKSAT--GIGLLISEGI 232
Db 181 QFVSVKASDVFLAVESYLLAKQIDQPHLGTTEAGARSAGVKSATFTGLLISEGI 240
QY 233 GDTLRVSLAADPVEEIKVGFIDILKSLRIRSRGINFIACPTCSROEPDVI--GTVNALEOR 290
Db 241 GDTLRVSLAADPVEEIKVGFIDILKSLRIRSRGINFIACPTCSROEPDVI--GTVNALEOR 300
QY 291 LEDITPMVDSIIGCVNPGGALVSTLGVGKNGKSGLYEDGVKRDRLD--NNDMIDOL 348
Db 301 LEDITPMVDSIIGCVNPGGALVSTLGVGKNGKSGLYEDGVKRDRLDFTNNDMIDOL 360
QY 349 EARIRAKASQLEARRIDVQOVER 372
Db 361 EARIRAKASQLEARRIDVQOVER 384

RESULT 3

US-08-827-190-5
Sequence 5, Application US/08827190
Patent No. 5858367
GENERAL INFORMATION:
APPLICANT: Rather, Philip N.
TITLE OF INVENTION: Methods For Screening For Antimicrobials
TITLE OF INVENTION: Utilizing aarc And Compositions Thereof
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Medlen & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/827,190
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: CASE-02443
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-827-190-5

Query Match 87.0%; Score 1622.5; DB 2; Length 365;
Best Local Similarity 88.4%; Pred. No. 1; 2e-168;
Matches 320; Conservative 24; Mismatches 17; Indels 1; Gaps 1;
QY 1 MHNQAPIQRKSTRIVGVNVPIDGAPIAVQSWTNTTRTDTVEATV--NQIKALERVGADI 60

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 17:14:33 ; Search time 21 Seconds
(without alignments)
749.507 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866

Sequence: 1 MHNQAPIORRSTRIYGVN.....RAKASQLEARRIDVOOVER 372

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_Aa:*

- 1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/prodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1800	96.5	384	2	US-08-827-190-6
2	1800	96.5	384	4	US-09-170-187-6
3	1622.5	87.0	365	2	US-08-827-190-5
4	1622.5	87.0	365	4	US-09-170-187-5
5	1372.5	73.6	547	4	US-09-252-991A-22991
6	1195	64.0	378	4	US-09-328-352-7906
7	470.5	25.2	621	4	US-09-198-452A-389
8	112.5	6.0	1058	4	US-09-252-991A-29105
9	111.5	6.0	578	3	US-09-066-046-6
10	106.5	5.7	980	4	US-09-252-991A-30838
11	105.5	5.7	578	3	US-08-975-762-50
12	105.5	5.7	578	3	US-09-295-028-50
13	105.5	5.7	578	3	US-09-106-582-50
14	101	5.4	465	4	US-09-252-991A-29387
15	101	5.4	523	4	US-09-612-964-2
16	100.5	5.4	620	4	US-09-328-352-7730
17	100	5.4	343	3	US-09-039-859-2
18	99.5	5.3	1374	4	US-09-252-991A-24636
19	97	5.2	1037	4	US-09-134-001C-4794
20	96.5	5.2	705	4	US-09-328-352-7436
21	96.5	5.2	851	4	US-09-252-991A-24773
22	95.5	5.1	303	4	US-09-134-001C-4855
23	95.5	5.1	740	1	US-08-309-512-10
24	95.5	5.1	740	5	PCT-US92-08756A-10
25	95	5.1	796	4	US-09-252-991A-17763
26	94.5	5.1	483	4	US-08-887-534A-51
27	94.5	5.1	483	4	US-09-527-431A-51

28	94.5	5.1	659	4	US-09-252-991A-17904	Sequence 17904, A
29	94.5	5.1	1503	3	US-08-976-255-14	Sequence 14, Appl
30	93.5	5.0	619	4	US-09-252-991A-17411	Sequence 17411, A
31	93.5	5.0	659	4	US-09-252-991A-17731	Sequence 17731, A
32	93	5.0	459	4	US-09-491-785-2	Sequence 2, Appl1
33	93	5.0	483	4	US-09-252-991A-27988	Sequence 27988, A
34	93	5.0	607	4	US-09-252-991A-20596	Sequence 20596, A
35	92.5	5.0	489	4	US-09-252-991A-27651	Sequence 27651, A
36	92.5	5.0	940	4	US-09-512-250C-2	Sequence 2, Appl1
37	91	4.9	470	4	US-09-252-991A-26312	Sequence 26312, A
38	91	4.9	817	4	US-09-252-991A-31147	Sequence 31147, A
39	91	4.9	1181	4	US-09-252-991A-18480	Sequence 18480, A
40	90.5	4.8	484	4	US-09-252-991A-20787	Sequence 20787, A
41	90.5	4.8	489	4	US-09-252-991A-19810	Sequence 19810, A
42	90.5	4.8	573	4	US-09-252-991A-26428	Sequence 26428, A
43	90.5	4.8	1512	1	US-08-169-927-2	Sequence 2, Appl1
44	90	4.8	358	4	US-09-784-508-4	Sequence 4, Appl1
45	90	4.8	629	4	US-09-252-991A-31575-	Sequence 31575, A

ALIGNMENTS

RESULT 1
US-08-827-190-6
; Sequence 6, Application US/08827190
; Patent No. 5858367
; GENERAL INFORMATION:
; APPLICANT: Rather, Philip N.
; TITLE OF INVENTION: Methods For Screening For Antimicrobials
; TITLE OF INVENTION: Utilizing aarc And Compositions Thereof
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Medtem & Carroll, LLP
; STREET: 220 Montgomery Street, Suite 2200
; CITY: San Francisco
; STATE: California
; COUNTRY: United States of America
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/827,190
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Carroll, Peter G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: CASE-02443
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 384 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-827-190-6
Query Match 96.5%; Score 1800; DB 2; Length 384;
Best Local Similarity 96.9%; Pred. No. 4.9e-188;
Matches 372; Conservative 0; Mismatches 0; Indels 12; Gaps 6;
QY 1 MHNQAPIORRSTRIYGVNPIGDCAPIAVOSMTNRTTVEATV-NQIKALEVGVADI 58
DB 1 MHNQAPIORRSTRIYGVNPIGDCAPIAVOSMTNRTTVEATVFTNQIKALEVGVADI 60
QY 59 VRVSTPTMDAAEAFILINQOVNVPVADIHFDYRIALAKVAEYGVDC--LRINPGINIGE 116

THIS PAGE BLANK (uspto)

DB 262 FEFAR1CRKLDYHNFVFSMKA5NPVIMVQAYRLVAVEMVHGWDYPLHLGVTGAGEGEGD 321
QY 217 AVKSAIGLLISEGIGDTLRVSLADPVEI 248
DB 322 RMKSAIGITLQDGLDITIRVSLTEPPEBEI 353

RESULT 15

08KG23
ID 08KG23 PRELIMINARY; PRT; 746 AA.
AC 08KG23;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE GCPE protein.
GN GCPE OR CT0147.
OS Chlorobium tepidum.
OC Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxId=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TLS / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=12093901;
RA Eisen J.A., Nelson K.B., Paulsen I.T., Heidelberg J.F., Wu M.,
Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
Nierman W.C., Feldblyum T.V., Hansen C.L., Craven M.B., Radune D.,
Vamathevan J., Khouli H., White O., Gruber T.M., Ketchum K.A.,
R.A. Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RT "The complete genome sequence of Chlorobium tepidum TLS, a
photoautotrophic, anaerobic, green-sulfur bacterium.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514(2002).
DR EMBL; AE012794; AAM71395.1; -.
DR TIGR; CT0147; -.
DR InterPro; IPR006705; GCPE.
DR InterPro; IPR004588; IspG.
DR Pfam; PF04551; GCPE; 2.
DR TIGRFAMs; TIGR00612; IspG_gcPE; 1.
KW Complete proteome.
SQ SEQUENCE 746 AA; 82087 MW; B40ED95B6010C5F9 CRC64;

Query Match 24.4%; Score 455.5; DB 16; Length 746;
Best Local Similarity 37.5%; Pred. No. 3.8e-25;
Matches 108; Conservative 51; Mismatches 90; Indels 39; Gaps 8;

QY 5 API---QRRKSTRIVYGVNPIGDGAPIAVQSMTRRTTVEATVNOIKALERYGADIVRV 61
DB 68 APVYSYRRRVTRVPPGTIFLGGYLPFRVSMITAHMTDPAASVEGCRRLYEGCEIIRL 127
QY 62 SVPTMDAAEAFKLIKQO---VNPPLVADIHFDYRIALKVAEYGVDCRLINPNTGN-- 114
DB 128 TVPTKDAENLNKIRIQRLRRDGDIDPLVADIHFSAKAMKAVEF--VENIRINPGNATGA 186
QY 115 -----EERIRMVVDGARDKNIRIRIGVNAVGSLEKDLQEKYGEPTPQ 155
DB 187 KRSSKDYTDDEYRAELDKVREBFTPLVRKRSIGVSMRIGTNGSLSDRIVSRYN-SPB 245
QY 156 ALLESAMRHVDHL-DRLNFDQFKSVKASDVFLAVESYRLAKOID-----QPLHLGIT 208
DB 246 GIVEAALERSRICEDGGYDQL-FSMKSSNVVRMIGAYRLVLVARADAEIRYAPLHLGVT 304
QY 209 EAGGARSGAVKSAIGLLISEGIGDTLRVSLADPVEEIKVGFDTLK 256
DB 305 EAGDGDGGRKSAIGLLISEGIGDTLRVSLTEDPVNEVPVGFALVK 352

Search completed: November 23, 2003, 17:16:38
Job time : 63 secs

DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
 DE Gcpe protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCBITaxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=97471969; PubMed=9330910;
 RA Sato S., Kocani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
 RA Miyajima N., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT clones."
 RL DNA Res. 4:215-230(1997).
 DR EMBL; AB005246; BAB09833.1; -;
 DR InterPro; IPR004588; IspG.
 DR TIGRFBMS; TIGR00612; IspG GCPE; 1.
 SQ SEQUENCE 716 AA; 79817 MW; 28D36FC64EB7CA CRC64;

Query Match 24.8%; Score 462; DB 10; Length 716;
 Best Local Similarity 38.6%; Pred. No. 1.2e-25;
 Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;

QY 9 RRSKTRIVYGNVPIGAGPIAVOSMTNRTTDEATVNOIKALERVADIVRSVPTMDA 68
 DB 59 RRTTRTVMGVNLGSEHPIRIQTMTSDTKDITGVDEVMRIADKAGIVITVOGKKE 118
 QY AEAFA-----KLTKQVNVPLVADIHFDYRIALKVAEYVDCRLINPNCINNEERIRMVVD 123
 DB 119 ADACFEIKKLVQNLVNNIPVADIHFPAPVALRVAC-PDKIRVNGNFPADRAQFETID 177
 QY 124 CARDK-----NIPRIGVNASGLEKDLQEKYGEPTQALLESA 161
 DB 178 YTEDEVQKELQIHIEQVFTPLVEKCKKYGARMRIGTNHGSLSDRIMSYGD-SPRGWVESA 236
 QY 162 MRHVHDLRLNFDQFKVSVKASDVFLAVESYRLAKOI-----DQPLHIGITEAGARG 216
 DB 237 FEFARICRLDYHNPFVSMKASNPVIMVQAYRLLVAMVYHGMVPLHLGVTENGEGEDG 296
 QY 217 AVKSAIGLILLEGISGIDTLRVSLAADPVEEI 248
 DB 297 RMKSAIGITLLODGLDGTIRVSLTEPPEEII 328

RESULT 13

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCBITaxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC Querol J., Campos N., Imperial S., Boronat A.,
 RA Rodriguez-Concepcion M.;
 RT "Identification and functional analysis of plant orthologs of the
 RT biochemistry. coli gcpE gene product with a role in plastid isoprenoid
 RT biosynthesis."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF344673; AA015446.1; -;
 DR TIGRFBMS; TIGR00612; IspG GCPE; 1.
 SQ SEQUENCE 740 AA; 82157 MW; 8EF625E9A9C88074 CRC64;

Query Match 24.8%; Score 462; DB 10; Length 740;

Best Local Similarity 38.6%; Pred. No. 1.2e-25;
 Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;

QY 9 RRSKTRIVYGNVPIGAGPIAVOSMTNRTTDEATVNOIKALERVADIVRSVPTMDA 68
 DB 84 RRTTRTVMGVNLGSEHPIRIQTMTSDTKDITGVDEVMRIADKAGIVITVOGKKE 143
 QY AEAFA-----KLTKQVNVPLVADIHFDYRIALKVAEYVDCRLINPNCINNEERIRMVVD 123
 DB 144 ADACFEIKKLVQNLVNNIPVADIHFPAPVALRVAC-PDKIRVNGNFPADRAQFETID 202
 QY 124 CARDK-----NIPRIGVNASGLEKDLQEKYGEPTQALLESA 161
 DB 203 YTEDEVQKELQIHIEQVFTPLVEKCKKYGARMRIGTNHGSLSDRIMSYGD-SPRGWVESA 261
 QY 162 MRHVHDLRLNFDQFKVSVKASDVFLAVESYRLAKOI-----DQPLHIGITEAGARG 216
 DB 262 FEFARICRLDYHNPFVSMKASNPVIMVQAYRLLVAMVYHGMVPLHLGVTENGEGEDG 321
 QY 217 AVKSAIGLILLEGISGIDTLRVSLAADPVEEI 248
 DB 322 RMKSAIGITLLODGLDGTIRVSLTEPPEEII 353

RESULT 14

OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 NC NCBITaxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
 RA Bowser L., Carlini P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Natusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RT Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY094472; AAM19840.1; -;
 DR InterPro; IPR006705; Gcpe.
 DR InterPro; IPR004588; IspG.
 DR Pfam; PF04551; Gcpe; 2.
 DR TIGRFBMS; TIGR00612; IspG GCPE; 1.
 SQ SEQUENCE 741 AA; 82257 MW; BCALD3147BD63ACB CRC64;

Query Match 24.5%; Score 457; DB 10; Length 741;
 Best Local Similarity 38.2%; Pred. No. 2.9e-25;
 Matches 104; Conservative 48; Mismatches 86; Indels 34; Gaps 5;

QY 9 RRSKTRIVYGNVPIGAGPIAVOSMTNRTTDEATVNOIKALERVADIVRSVPTMDA 68
 DB 84 RRTTRTVMGVNLGSEHPIRIQTMTSDTKDITGVDEVMRIADKAGIVITVOGKKE 143
 QY AEAFA-----KLTKQVNVPLVADIHFDYRIALKVAEYVDCRLINPNCINNEERIRMVVD 123
 DB 144 ADACFEIKKLVQNLVNNIPVADIHFPAPVALRVAC-PDKIRVNGNFPADRAQFETID 202
 QY 124 CARDK-----NIPRIGVNASGLEKDLQEKYGEPTQALLESA 161
 DB 203 YTEDEVQKELQIHIEQVFTPLVEKCKKYGARMRIGTNHGSLSDRIMSYGD-SPRGWVESA 261
 QY 162 MRHVHDLRLNFDQFKVSVKASDVFLAVESYRLAKOI-----DQPLHIGITEAGARG 216

DR EMBL; A014468; AAN30676.1; --
 DR TIGR; BR178; --
 KW Complete proteome.
 SQ SEQUENCE 420 AA; 45029 MW; F1495DB2D73E164C CRC64;

Query Match 27.2%; Score 508; DB 16; Length 420;
 Best Local Similarity 33.2%; Pred. No. 2.2e-29;
 Matches 128; Conservative 79; Mismatches 109; Indels 70; Gaps 13;

QY 6 PIRKSTRIVGNVPIDGAPVAGSMNTRTTVDVATVNOIKALERVGADIVRSVPT 65
 DB 12 PFRFROSGVGVSGVIVGSAFVVOVMNTIDTADVSTVAGVAAHRAAGSEIVRTVDR 71
 QY 66 MDAEAFKLIQOV-----NVPLVADIF-----DY-RIALKVAEYVDCIRINPNT 112
 DB 72 DESAAAVPKIRERLERGHVPLVGPDPHYIGHKLADHPKAEALAKY-----RINPNT 126
 QY 113 G---NEERIRMVVDCRDKNIPIRIGVNAAGLEKDI-----OEKYGEP-----Q 155
 DB 127 GFKDKDKOPADIVEMAIRYDKERVIGVNMGSLDOELLTTIMDRNMQEAPLSAODVMRE 186
 QY 156 ALLESAMRHVDHLDRLNF--DQKVSVKASDVFLAVESYLLAKQIDQPLHITAGGA 213
 DB 187 AIVQSLALISANLAEIIGLRDKITLSAKVSQVODLIAVYMLQKRSVHALHLTLTGWGC 246
 QY 214 RSGAVKSAIGLGLLSEIGDITLRVSLADP---VEEIKVGFPIKSLIRSRGINFLA 269
 DB 247 TKGIVASAMGILLQOIGDITIRISLTPRGGDRREVQVAGELQITMGFRQFVIVA 306
 QY 270 CPTCSQOEPFVITVNALEQRL-----EDI-----ITPVSITICVNVPG 311
 DB 307 CFCGCR-----TTSVFQELAQITQIEDIRRMPLREKYPGEVALSVVMGCIYVPG 359
 QY 312 EALVSTLGV-----TGKTKSGLYEDG 333
 DB 360 ESKHADIGISLPGETPSPAPVVDG 385

RESULT 10

Q8GZR6 PRELIMINARY; PRT; 740 AA.
 AC Q8GZR6;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE GCPE.
 OS Lycopersicon esculentum (Tomato).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; lamiales; Solanales; Solanaceae; Solanum.
 OC NCBI_Taxid=4081;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Querol J., Imperial S., Boronat A., Rodriguez-Concepcion M.;
 RT "Identification and functional analysis of plant orthologs of the
 RT Escherichia coli gcpe gene product with a role in plastid isoprenoid
 RT biosynthesis."
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF435086; AA015447.1; --
 SQ SEQUENCE 740 AA; 82230 MW; 8668C9F08078FA05 CRC64;

Query Match 25.2%; Score 471; DB 10; Length 740;
 Best Local Similarity 40.1%; Pred. No. 2.7e-26;
 Matches 109; Conservative 45; Mismatches 84; Indels 34; Gaps 5;

QY 9 RRSKSTRIVGNVPIDGAPVAGSMNTRTTVDVATVNOIKALERVGADIVRSVPTMDA 68
 DB 84 RRGTCVTVMGVALGSEHPIRIGTMTTDDVADAVEGQMKADAGADIVRTVQGRKE 143
 QY 69 AEAFF-----KLKQOVNPLVADIHFDYRIALKVAEYVDCIRINPNTGN----- 114
 DB 144 ADACFEIKQTLVQKNYINPLVADIHFAFPAVALRAEC-FDKIRVNGNPNADRRAPPEOLE 202

QY 115 -----EERIRNVDCARDKNIPIRIGVNAAGLEKDLQEKYGEPTPOALLESA 161
 DB 203 YTEDDYQKELEHIEBEVFTPLVEKCKYGRAMRIGTHGSLSRIMSYGD-SPRGWESA 261
 QY 162 MRHVHDLRLNFDQFVSVKASDVFLAVESYRLAKQI-----DQPLHITAGGARSG 216
 DB 262 FEFARICRLDHFNFVFSMKASNPVMVOAYRLVAEMYVQGWDPYHLHGVTEAGGEDG 321
 QY 217 AVKSAIGLGLLSEIGDITLRVSLADPVEEI 248
 DB 322 RMXSAIGITGLDQDGLDITIRVSLTEPPEEI 353

RESULT 11

Q8RXG8 PRELIMINARY; PRT; 741 AA.
 AC Q8RXG8;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE GCPE protein.
 GN AT5G60600.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC euroside II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_Taxid=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carinici P., Chen H.,
 RA Cherk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY081261; AAL91150.1; --
 DR InterPro; IPR006705; GCPE.
 DR InterPro; IPR004588; IPGC.
 DR Pfam; PF04551; GCPE; 1.
 DR TIGRPFAM; TIGR00612; IPGC_gcpe; 1.
 SQ SEQUENCE 741 AA; 82157 MW; 957B11538F481585 CRC64;

Query Match 24.8%; Score 463; DB 10; Length 741;
 Best Local Similarity 38.6%; Pred. No. 1e-25;
 Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;

QY 9 RRSKSTRIVGNVPIDGAPVAGSMNTRTTVDVATVNOIKALERVGADIVRSVPTMDA 68
 DB 84 RKRTRVVMGVALGSEHPIRIGTMTTSDTKDITGVDEVMRIADKADIVRTVQGRKE 143
 QY 69 AEAFF-----KLKQOVNPLVADIHFDYRIALKVAEYVDCIRINPNTGNBEERIRNV 123
 DB 144 ADACFEIKQTLVQKNYINPLVADIHFAFPAVALRAEC-FDKIRVNGNPNADRRAPPEETD 202
 QY 124 CARDK-----NIPRIGVNAAGLEKDLQEKYGEPTPOALLESA 161
 DB 203 YTEDEYQKELEHIEVFTPLVEKCKYGRAMRIGTHGSLSRIMSYGD-SPRGWESA 261
 QY 162 MRHVHDLRLNFDQFVSVKASDVFLAVESYRLAKQI-----DQPLHITAGGARSG 216
 DB 262 FEFARICRLDYNHNFVFSMKASNPVMVOAYRLVAEMYVHGMWDYRLHGVTEAGGEDG 321
 QY 217 AVKSAIGLGLLSEIGDITLRVSLADPVEEI 248
 DB 322 RMXSAIGITGLDQDGLDITIRVSLTEPPEEI 353

RESULT 12

Q9FF59 PRELIMINARY; PRT; 716 AA.
 AC Q9FF59;
 DT 01-MAR-2001 (TReMBLrel. 16, Created)

```

QY 189 VESYRLAKOIQDPLHIGITEAGASGANKSAIGLGLLSEIGDITLRSVLAADPVEEL 248
DB 185 IEAVYKAAAEFNFYHLHIGITEAGSLFENGITKSAAGVLLHIEGINTIRISLIDPLESEV 244
QY 249 KVFQDLKSLRISRSRGINFIACPTCSROEPDVIGTVAALEQRLEDDITPMDVSIICGVN 308
DB 245 KVCCKLINSIGLVNDLVDIISCTCGRLNFDLNPVKEIKFRKRNFKPLKVAIIGCAN 304
QY 309 GPGCALVSTLGVTVGNKKSGLYEDGVKRDLDNNMDIDLEARIKAKASQLEARR 364
DB 305 GPGEAKEADIGIAGNGTGIIFANGRKAIKSVPEQDVLDELKLIISIKYKEYLDSKK 360

RESULT 7
Q8DK70 PRELIMINARY; PRT; 402 AA.
Q8DK70
AC Q8DK70;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE TLR0996.
GN TLR0996.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_Taxid=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneke T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Igaruchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno M., Nakazaki N.,
RA Shimpo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002)
KM EMBL; AP005372; BAC08548.1; -
KW Complete proteome.
SQ SEQUENCE 402 AA; 44192 MW; 7595FFC940AC60D7 CRC64;

Query Match 32.2%; Score 601.5; DB 16; Length 402;
Best local Similarity 36.8%; Pred. No. 2,4e-36;
Matches 147; Conservative 79; Mismatches 120; Indels 53; Gaps 11;

QY 4 QADIQRRKSTRIVYGNVPIGDGAPIAVQSMNTRTTVDVATNQIKALERVGADIVRSV 63
DB 14 ETATVRRKTRPVPIGVSIVIGGHPVAQVMINEDTLDEGSVAIRRLHEIGCEIVATV 73
QY 64 PTMDAAEAF-----KLIKQVNPVLVDIHF-D-YRIKLKVAEYGVDCLRINPG----- 110
DB 74 PSLAAKAMEIRDLRYKTPVPLVDVHNGKIALEVAKY-VDVVRINPGIYVEKP 132
QY 111 -----NIGNE--ERIMVVDCAADKNIPIRIGVNGSLSEKDOEKYGEPTPQ 155
DB 133 KPNRTETQAFPEDEIGAKIETLEPLVISLRDQCKSMRIGVNGSLAERLFTYGD-TPE 191
QY 156 ALLESAMRHVDHLDRINFDFQFKVSKASDVFLAVESYRLAKOIQD-----PLHLGITEA 210
DB 192 GWESALEFIRICSLNFYMLISILKASRVPMIAANRLMKRDELGMOPYPLHLGITEA 251
QY 211 GGAASGAVKSAIGLGLLSEIGDITLRSVLAADPVEEIKYGFDLKSLRSGINFIAC 270
DB 252 GDGEYGIKSTAGIATLARIIGITIVSLTEAPEKEIPVCGYLLQGLRRTWVEVAC 311
QY 271 PTCRSRQFDVIGTVNALEQRLEDDITPMDVSIICGVNPGCALVSTLGVTVGN 324
DB 312 PSCGRTIFN-----LEEVIAHKREARKILTGINIAVMGIVNGPEMADADGYGV-- 362
QY 325 KKSQ--LYEDGVKRDLDNNMDIDLEARIKAKASQLE 360
DB 363 KOPGYISLYRGREVKVPEAEVAAVLEIKADGRWVD 401

```

```

RESULT 8
Q8FLH5 PRELIMINARY; PRT; 663 AA.
Q8FLH5
AC Q8FLH5;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE GCPE protein homolog.
GN GCPE OR LA3160.
OS Leptospira interrogans.
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
OX NCBI_Taxid=173;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RA Ren S.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE011478; AAN50358.1; -
KW Complete proteome.
SQ SEQUENCE 663 AA; 73779 MW; 79B60DF805A6365 CRC64;

Query Match 28.2%; Score 525.5; DB 16; Length 663;
Best local Similarity 41.7%; Pred. No. 2.1e-30;
Matches 118; Conservative 53; Mismatches 75; Indels 37; Gaps 8;

QY 2 HNOAPI--QRKSTRIVYGNVPIGDGAPIAVQSMNTRTTVDVATNQIKALERVGADIV 59
DB 5 YNOTPEGYORRRKTRPVPIGVSIVIGGHPVAQVMINEDTLDEGSVAIRRLHEIGCEIV 64
QY 60 RVSVPMDAAEAFKLIKQOV-----NVPLVADIHFPYRIALKAEGYVDCLRINPGIN 114
DB 65 RLTVESQADADLPSIRQLKKAAGSVPLVDIHFPPSAMKAVEI-VEKVRINPGNFD 123
QY 115 E-----BRIR-----MVVDCARDKNIPRIGVNGSLSEKDOEKYGEPTPQ 152
DB 124 KKKFAVRDYTDLEVNGELERISEVFPLVIRC-KELGVSMRIGTNGHSLSDRIMRYGD- 181
QY 153 TPQALLSARHVDHLDRINFDFQFKVSKASDVFLAVESYRLAK-----QIDQPLHLGI 207
DB 182 TPQWESALEFIRIAESLGYVDIIVSMKASNPQVQAYRMLASFNELKMDYPLHLGV 241
QY 208 TEAGASGAVKSAIGLGLLSEIGDITLRSVLAADPVEEIKV 250
DB 242 TEAGDNGDRIRKSAIGSLDEGLDITRVSITBEDVLEVEV 264

RESULT 9
Q8FYT2 PRELIMINARY; PRT; 420 AA.
Q8FYT2
AC Q8FYT2;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE GCPE protein.
GN GCPE OR BR1778.
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_Taxid=29461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122;
RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
RA Daugherty S.C., Debroy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
RA Riedmiller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,
RA Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M., Frazer C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts.";
Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).

```


ISFG ORBL0098.

OS Bifidobacterium longum.
OC Bacteria; Actinomycetia; Actinobacteridae; Bifidobacterales;
XC Bifidobacteriaceae; Bifidobacterium.
OX NCBI_TaxID=216816;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCC 2705;
RX MEDLINE=22294977; PubMed=12381787;
RA Schell M.A., Karmaliantzou M., Snel B., Vilanova D., Berger B.,
RA Pessi G., Zwielen W.-C., Desiere F., Bork P., Delley M.,
RA Pidmore R.D., Arigoni F.;
RT "The genome sequence of Bifidobacterium longum reflects its adaptation
to the human gastrointestinal tract." ;
RL Proc. Natl. Acad. Sci. U.S.A. 99:14422-14427 (2002).
DR EMBL; AE014624; AAN23963.1; -.
KM Complete proteome.
SQ SEQUENCE 403 AA; 42564 MW; FF68CAJ6153A6056 CRC64;

Query Match 43.6%; Score 814.5; DB 16; Length 403;
Best Local Similarity 46.3%; Pred. No. 4.2e-52;
Matches 167; Conservative 70; Mismatches 111; Indels 13; Gaps 4

DQ 3 NOADPQ-RKSGTRIYGVNVPDGDGAPIVAGSTWNTTDTVDVATVOIKALRGVADIYRV 61
 ::::| | | | | | | | | | | | | | | | | : | | | | | | | | | |
Db 22 SESFHPHRRKRRIWGPVVGGGAPISVSQTWNLTLMNPPTLOOIMELTPACDIYRV 81

DY 62 SVPTMDAAEAEXLIKQOVNVPLVVDIHFDYSIALKAVEYGVCLEIRINPGINEERIRM- 120
 :::|| ||| | | : | : | : | | | | | | | : | : | | | | : | :
DB 82 AVPSQGDDDALPEICRKSPIVIADIIHQSKVFPQIADGCAGAARVNNGNIKKPFEPVDP 141

DY 121 VVDCARDKNIPDIRIGVNAAGSLSEKDIOEKYEETPOLLESAMRHVDHLDRINFQFKSV 180
 142 ICKATPDAGISLRIGVNAGSLDKELYAKYCGPTPEALTVAALKKEAMPEDVGDFHFXTS 201

DY 181 KASVPFLAVESYRLLANKQIQDPLEHGTEAGARGASKSAIGCLLSSEGIGDTLTRVL 240
 ||||| | | | | | | | | | | | | | | | | : | : | | | | | | | |
DB 202 KHGHVTITWEHYRYRLASKGDMPLHGVTAEAPAWOGTTKSCIAEGALLAEGI GTTRVSL 261

DY 241 AADVEBEIKVFPDFILKSLIRSRCGINFIACPCTCSROEFDVGTVALREQLEDITTPMDV 300
 STFAIN=Y\$-314 / Yamasaki J., Hino Y., Kikuchi H., Nakamura Y.,
DB 262 SAPPAEEVKYCCKLEVMGLPRPKREDITSCHSCGAQADVQLASAVEGLEKDYAPAIRV 321

DY 301 SIICVVANGCEPALVSTLVGYTGNNKGSGLYBDG----VRQRLDN-----NDMIDDLE 349
 ::::| | | | | | | | | | | | | | | | | : | : | | | | | | | |
DB 322 AMGCCIVGPCGEAREADDLGVASGNKGQIFTKGVKITVPEDQIVDTLLTIANDIAQME 381

DY 350 A 350
 Matches 167; Conservative 70; Mismatches 111; Indels 13; Gaps 4
DB 382 A 382

RESULT 5

OQBP82 PRELIMITARY; PRT: 393 AA.

AC OQBP82;
DB 01-MAR-2003 (TREMBLrel_23, Created)
DT 01-MAR-2003 (TREMBLrel_23, Last sequence update)
DE 01-MAR-2003 (TREMBLrel_23, last annotation update)
DN Putative aminoglycoside acetyltransferase negative regulator.
CN CEI903.
OC Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
CX Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y\$-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeo K., Suzuki M., Maehima J., Itoh T., Yamagishi A., Nishio Y.,
RA Ueda Y., Sugimoto S.;
RT The entire genomic sequence of Corynebacterium efficiens Y\$-314.",
Submitted (May-2002) to the EMBL/Genebank/DBJ databases.

Query Match	40.4%	Score 753	DB 16	Length 393
Best Local Similarity	43.7%	Pred. No. 1.4e-47		
Matches 157	Conservative	67	Mismatches 123	Indels 12
				Gaps 3

ID	Q8EUT6	PRELIMINARY	PRT	362 AA
AC	Q8EUT6			
DT	01-MAR-2003	(TREMBLrel. 23, Created)		
DT	01-MAR-2003	(TREMBLrel. 23, Last sequence update)		
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)		
DE	Peptidoglycan acetylation.			
GN	MYB9400.			
OS	Mycoplasma penetrans.			
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.			
OX	NCBI_TaxID=28227;			
RX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=HF-2;			
RX	MEDLINE=22354719; PubMed=12466555;			
RA	Sasaki Y., Ishikawa J., Yamaashita A., Oshima K., Kenri T., Furuya K., Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.,			
RT	"The complete genomic sequence of Mycoplasma penetrans, an intracellular bacterial pathogen in humans."			
RL	Nucleic Acids Res. 30:5293-5300(2002).			
DR	EMBL; AP004174; BAC44727.1; -.			
KM	Complete proteome.			
SR	SEQUENCE 362 AA; 39237 MW; 3F7EB12EF331AE73 CRC64;			

Query Match	39.6%	Score 739.5	DB 16	Length 362
Best Local Similarity	41.0%	Pred. No. 1.3e-46		
Matches 146	Conservative	82	Mismatches 127	Indels 1
				Gaps 1

ID	Q8EUT6	PRELIMINARY	PRT	362 AA
AC	Q8EUT6			
DT	01-MAR-2003	(TREMBLrel. 23, Created)		
DT	01-MAR-2003	(TREMBLrel. 23, Last sequence update)		
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)		
DE	Peptidoglycan acetylation.			
GN	MYB9400.			
OS	Mycoplasma penetrans.			
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.			
OX	NCBI_TaxID=28227;			
RX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=HF-2;			
RX	MEDLINE=22354719; PubMed=12466555;			
RA	Sasaki Y., Ishikawa J., Yamaashita A., Oshima K., Kenri T., Furuya K., Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.,			
RT	"The complete genomic sequence of Mycoplasma penetrans, an intracellular bacterial pathogen in humans."			
RL	Nucleic Acids Res. 30:5293-5300(2002).			
DR	EMBL; AP004174; BAC44727.1; -.			
KM	Complete proteome.			
SR	SEQUENCE 362 AA; 39237 MW; 3F7EB12EF331AE73 CRC64;			

Query Match	40.4%	Score 753	DB 16	Length 393
Best Local Similarity	43.7%	Pred. No. 1.4e-47		
Matches 157	Conservative	67	Mismatches 123	Indels 12
				Gaps 3

ID	Q8EUT6	PRELIMINARY	PRT	362 AA
AC	Q8EUT6			
DT	01-MAR-2003	(TREMBLrel. 23, Created)		
DT	01-MAR-2003	(TREMBLrel. 23, Last sequence update)		
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)		
DE	Peptidoglycan acetylation.			
GN	MYB9400.			
OS	Mycoplasma penetrans.			
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.			
OX	NCBI_TaxID=28227;			
RX	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=HF-2;			
RX	MEDLINE=22354719; PubMed=12466555;			
RA	Sasaki Y., Ishikawa J., Yamaashita A., Oshima K., Kenri T., Furuya K., Yoshino C., Horino A., Shiba T., Sasaki T., Hattori M.,			
RT	"The complete genomic sequence of Mycoplasma penetrans, an intracellular bacterial pathogen in humans."			
RL	Nucleic Acids Res. 30:5293-5300(2002).			
DR	EMBL; AP004174; BAC44727.1; -.			
KM	Complete proteome.			
SR	SEQUENCE 362 AA; 39237 MW; 3F7EB12EF331AE73 CRC64;			

Query Match	40.4%	Score 753	DB 16	Length 393
Best Local Similarity	43.7%	Pred. No. 1.4e-47		
Matches 157	Conservative	67	Mismatches 123	Indels 12
				Gaps 3

ID	Q8EUT6	PRELIMINARY	PRT	362 AA
AC	Q8EUT6			
DT	01-MAR-2003	(TREMBLrel. 23, Created)		
DT	01-MAR-2003	(TREMBLrel. 23, Last sequence update)		
DT	01-MAR-2003	(TREMBLrel. 23, Last annotation update)		
DE	Peptidoglycan acetylation.			
GN	MYB9400.			
OS	Mycoplasma penetrans.			
OC	Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.			
OX	NCBI_TaxID=28227;			
RX	[1]			
RP				

QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSSGAVKSAIGLISGIGDTLRVSL 240
 DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSSGAVKSAIGLISGIGDTLRVSL 240
 QY 241 AADPVEIKVGFDLKSLRIRSRGINFIACPTCSROEFDVIGTVALERLEDITPMDV 300
 DB 241 AADPVEIKVGFDLKSLRIRSRGINFIACPTCSROEFDVIGTVALERLEDITPMDV 300
 QY 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKASQ 359
 DB 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKASQ 359
 QY 360 DEARRIDVQ 368
 DB 361 DESNRIDIK 369

RESULT 2

Q8EC32 PRELIMINARY; PRT: 371 AA.
 AC Q8EC32;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
 GN ISP6 OR SO3312.
 OS Shewanella oneidensis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Alteromonadales; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NR-1;
 RX MEDLINE=22297686; PubMed=12368813;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 RA Read T.D., Eisen J.A., Seehadri R., Ward N., Methe B., Clayton R.A.,
 RA Meyer T., Tappin A., Scott J., Beanan M., Brinkac L., Dougherty S.,
 RA Deboy R.T., Dodson R.J., Durkin A.S., Haft D.H., Kolonay J.F.,
 RA Madupu R., Peterson J.D., Umayam L.A., White O., Wolf A.M.,
 RA Vamathevan J., Weidman J., Impraim M., Lee K., Berry K., Lee C.,
 RA Mueller J., Khouri H., Gill J., Utecherback T.R., McDonald L.A.,
 RA Feldblyum T.V., Smith H.O., Venter J.C., Neilson K.H., Fraser C.M.;
 RT "genome sequence of the dissimilatory metal ion-reducing bacterium
 RT Shewanella oneidensis";
 RL Nat. Biotechnol. 20:1118-1123(2002).
 DR EMBL; AB015769; AAN56310.1; -;
 DR TIGR; SO3312; -;
 KM Complete proteome.
 SQ SEQUENCE 371 AA; 40603 MW; 024A994C6EC93906 CRC64;

Query Match 84.9%; Score 1584.5; DB 16; Length 371;
 Best Local Similarity 84.8%; Pred. No. 3.8e-109;
 Matches 313; Conservative 32; Mismatches 21; Indels 3; Gaps 2;

QY 1 MANOAPLQRRKSTRITVYGNVPIGDAPIAVQSMNTTDTTVEATVNOIKALERGADIVR 60
 DB 1 MYNETPIKRBPSTRTIYGNVPIGDAPIAVQSMNTTDTTVEATVNOIKALERGADIVR 60
 QY 61 VSVPTMAAFAFKLIKQOVNVPVADIHFDYRIALKAAYGVDCRLRNPGNIGNEERIRM 120
 DB 61 VSVPTMAAFAFKLIKQOVNVPVADIHFDYRIALKAAYGVDCRLRNPGNIGNEERIRM 120
 QY 121 VVDCARDKNIPIRIGVAGSLKDLQEKYGEPTPOALLSARHVDHLRLNFDQFVSV 180
 DB 121 VVDCARDKNIPIRIGVAGSLKDLQEKYGEPTPOALLSARHVDHLRLNFDQFVSV 180
 QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSSGAVKSAIGLISGIGDTLRVSL 240
 DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSSGAVKSAIGLISGIGDTLRVSL 240
 QY 241 AADPVEIKVGFDLKSLRIRSRGINFIACPTCSROEFDVIGTVALERLEDITPMDV 300
 DB 241 AADPVEIKVGFDLKSLRIRSRGINFIACPTCSROEFDVIGTVALERLEDITPMDV 300

DB 241 AADPVEIKVGFDLKSLRIRSRGINFIACPTCSROEFDVIGTVALERLEDITPMDV 300
 QY 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKASQ 359
 DB 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKASQ 359
 QY 360 DEARRIDVQ 368
 DB 361 --ANRIQVK 367

RESULT 3

Q8DIY3 PRELIMINARY; PRT: 366 AA.
 AC Q8DIY3;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE GCPE protein.
 GN GCPE
 OS Wigglesworthia brevipalpis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Wigglesworthia.
 OX NCBI_TaxID=164609;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22297718; PubMed=12219091;
 RA Akman L., Yamashita A., Watanabe H., Oshima K., Shiba T., Hattori M.,
 RA Aksoy S.;
 RT "Genome sequence of the endocellular obligate symbiont of tsetse
 RT flies, Wigglesworthia glosiniidida";
 RL Nat. Genet. 32:402-407(2002).
 DR EMBL; AB063522; BAC24719.1; -;
 KM Complete proteome.
 SQ SEQUENCE 366 AA; 40407 MW; 159E0169B8FC8DED CRC64;

Query Match 65.7%; Score 1226.5; DB 16; Length 366;
 Best Local Similarity 65.9%; Pred. No. 1.2e-82;
 Matches 234; Conservative 62; Mismatches 58; Indels 1; Gaps 1;

QY 7 IQRRKSTRITVYGNVPIGDAPIAVQSMNTTDTTVEATVNOIKALERGADIVR 66
 DB 7 IIRRSKSTIKYIKVPIGDSPISSVQSMNTTCTDINSTISQINKQKAGADIVR 66
 QY 67 DAAEAFKLIKQOVNVPVADIHFDYRIALKAAYGVDCRLRNPGNIGNEERIRMVDCAR 126
 DB 67 EAAEAFKLIKQOVNVPVADIHFDYRIALKAAYGVDCRLRNPGNIGNIIISVSTAK 126
 QY 127 DKNIPIRIGVAGSLKDLQEKYGEPTPOALLSARHVDHLRLNFDQFVSVKASDVR 186
 DB 127 EKKLPRIIGVAGSLKDLQEKYGEPTPOALLSARHVDHLRLNFDQFVSVKASDVR 186
 QY 187 LAVESYRLAKQIDOPHLGITEAGARSSGAVKSAIGLISGIGDTLRVSLADPVE 246
 DB 187 TCVQSKYKLAASKIDQDLHGLITESGMLHSGISIGLISGIGDTLRVSLADPVE 246
 QY 247 EIVYGVDFLKSRLIRSRGINFIACPTCSROEFDVIGTVALERLEDITPMDVSIIGCV 306
 DB 247 EIVYGVDFLKSRLIRSRGINFIACPTCSROEFDVIGTVALERLEDITPMDVSIIGCV 306
 QY 307 VNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKASQ 360
 DB 307 VNGPGEALVSTLGTGNGKSGLYEDGVR-KDRLDNDMDIDLEARIKASQ 360

RESULT 4

Q8G7Y6 PRELIMINARY; PRT: 403 AA.
 AC Q8G7Y6;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase 1.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 17:02:08 ; Search time 59 Seconds
(without alignments)
1627.044 Million cell updates/sec

Title: US-09-921-992-50
Perfect score: 1666
Sequence: 1 MHNQAPIQRKSTRIVGVN.....RAKASQDEARRIDVQVEK 372

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1631.5	87.4	372	16	Q8DEZ8 vibrio vuln
2	1584.5	84.9	371	16	Q8EC32 shewanella
3	1226.5	65.7	366	16	Q8DIY3 wigglisworth
4	814.5	43.6	403	16	Q8G7Y6 bifidobacte
5	753	40.4	393	16	Q8FP82 corynebacte
6	739.5	39.6	362	16	Q8EUI6 mycoplasma
7	601.5	32.2	402	16	Q8DK70 synecococc
8	525.5	28.2	663	16	Q8FTH5 leptospira
9	508	27.2	420	16	Q8FYT2 brucella su
10	471	25.2	740	10	Q8GZK6 lycopersico
11	463	24.8	741	10	Q8RGY8 arabidopsis
12	462	24.8	716	10	Q8PFS9 arabidopsis
13	462	24.8	740	10	Q8GZK7 arabidopsis
14	457	24.5	741	10	Q8LPO4 arabidopsis
15	455.5	24.4	746	16	Q8KG23 chlorobium
16	446	23.9	824	5	Q9BJX5 plasmodium

17	446	23.9	824	5	Q8IUH7	Q8IUH7 plasmodium
18	136	7.3	803	17	Q8R927	Q8R927 thermoaer
19	122.5	6.6	507	17	Q8R2P6	Q8R2P6 pyrococcus
20	119	6.4	535	16	Q8D332	Q8D332 synecococc
21	114.5	6.1	588	2	Q8VS09	Q8VS09 actinobacil
22	114	6.1	524	17	Q27773	Q27773 methanobact
23	112.5	6.0	686	17	Q9UXG1	Q9UXG1 sulfobolus
24	112.5	6.0	723	3	Q9PAD5	Q9PAD5 emericella
25	111.5	6.0	344	16	Q8RQA9	Q8RQA9 oceanobacil
26	111.5	6.0	578	2	Q68221	Q68221 anaplasmia p
27	111.5	6.0	588	2	Q8VSR0	Q8VSR0 actinobacil
28	111	5.9	755	16	Q8XYK5	Q8XYK5 ralestonia s
29	109	5.8	742	2	Q9RNG5	Q9RNG5 leptonella
30	108.5	5.8	449	16	Q8PZ29	Q8PZ29 xanthomonas
31	108.5	5.8	1000	17	Q8RPA1	Q8RPA1 methanobact
32	106.5	5.7	284	16	Q9CHV6	Q9CHV6 methanobact
33	106.5	5.7	798	2	Q9KHS7	Q9KHS7 pseudomonas
34	106.5	5.7	798	2	Q8KURS	Q8KURS pseudomonas
35	106	5.7	748	17	Q57616	Q57616 methanococc
36	105.5	5.6	578	2	Q93MY9	Q93MY9 anaplasmia p
37	104.5	5.6	638	5	Q9BMQ6	Q9BMQ6 opisthorchi
38	104	5.6	606	16	Q8Y7W3	Q8Y7W3 listeria mo
39	104	5.6	1227	16	Q97K41	Q97K41 clostridium
40	103.5	5.5	537	16	Q8FL75	Q8FL75 escherichia
41	102.5	5.5	417	3	Q9CA73	Q9CA73 emericella
42	102	5.5	546	16	Q8D263	Q8D263 streptococc
43	102	5.5	546	16	Q8DX35	Q8DX35 streptococc
44	101.5	5.4	564	5	Q44422	Q44422 heliothis v
45	101.5	5.4	572	10	Q8GZY8	Q8GZY8 oryza sativ

ALIGNMENTS

RESULT 1

Q8DEZ8 PRELIMINARY; PRT; 372 AA.

AC Q8DEZ8;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DE Gcpe.
GN VV10427.
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CMCP6;
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,
RA Choy H.E.;
RT "Complete genome sequence of Vibrio vulnificus CMCP6";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE016798; AAC008950.1; -;
KW Complete proteome.
SQ SEQUENCE 372 AA; 40573 MM; 6C73C2B8819B0285 CRC64;

Query Match 87.4%; Score 1631.5; DB 16; Length 372;
Best Local Similarity 86.7%; Pred. No. 1.3e-112;
Matches 320; Conservative 29; Mismatches 19; Indels 1; Gaps 1;

QY	1	MHNQAPIQRKSTRIVGVNPIGDGAPVAVQSMTRTTDVEATVNOIKALEVGDIVR 60
DB	1	MHNESPIIRKSTRIVGVNPIGDGAPVAVQSMTRTTDVAATVAQIKALEVGDIVR 60
QY	61	VSVPTMDAAEAPFLIKQOVNPLVADIHDPYRATLVAEYGVDCILINPGENEERIRM 120
DB	61	VSVPTMDAAEAPFLIKQOVNPLVADIHDPYRATLVAEYGVDCILINPGENEERIRN 120
QY	121	VVDCARDKNIPRIGVAGSLKDLQKYEPTPQALLSARHVRHDLRLNDFQKVS 180
DB	121	VVDCARDKNIPRIGVAGSLKDLQKYEPTPQALLSARHVRHDLRLNDFQKVS 180

RESULT 15
ISPG CAUCR STANDARD; PRT; 383 AA.
AC 09A9W0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ISPG OR CC0851.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacterales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nieman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson R.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Hart D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Ueberback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus";
Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
-1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
(ME-2,4CP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
(By similarity).
CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE005761; AKR2836.1; -
DR PIR; H87354; H87354.
DR TIGR; CC0851; -
DR HAMAP; MF_00159; -; 1.
DR InterPro; IPR004588; ISPG.
DR Pfam; PF04551; GcPB; 1.
DR TIGRFAMs; TIGR00612; ISPG_gcPB; 1.
KM Isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 383 AA; 40788 MW; 8468EB400B38194 CRC64;

Query Match 55.9%; Score 1042.5; DB 1; Length 383;
Best Local Similarity 57.6%; Pred. No. 9.2e-65;
Matches 208; Conservative 57; Mismatches 95; Indels 1; Gaps 1;

QY 7 IORRKRTRIVGVNPIGDGPIAVQSMTRTRTDVEATVNOIKALEVGDIVRVSVPTM 66
DB 14 ITRRQSRKIRVSGVEVGDAPISVQSMNTNLTSDAATLTLEIQLEAGADIVRVSCPDV 73
QY 67 DAAEAFKLIRKQVNVPLVADIHEDYRITALKVAEYGVDCLRINPGINERIRRVVDCAR 126
DB 74 ESTRAFKTIRAEKVPPLVADIHEDYRITALKVAEYGVDCLRINPGINERIRRVVDCAR 133
QY 127 DKNIPIRIGVNAAGLEKDLQEKYGEPTPOALLBSAMRHVDHLDRLNDQFKVSVKASDV 186
DB 134 DHCCSMRIGVNAAGLEKDLQEKYGEPTPOALLBSAMRHVDHLDRLNDQFKVSVKASDV 193
QY 187 LAVESYLLAKQIDQPHLITENGARSGAVKATGLGILLSGIDTILRVSLADPVE 246
DB 194 MTAAYAYQLAEVADICPHLITENGARSGAVKATGLGILLSGIDTILRVSLADPVE 253
QY 247 EIKVGFILKSLIRSRGINFIACPTCSRQEFVIGTVNALQGLIEDIITPMVSIIGCV 306

DB 254 EIKVGFILKSLIRSRGINFIACPTCSRQEFVIGTVNALQGLIEDIITPMVSIIGCV 313
QY 307 VNGPGALVSTIGVNGKNGSGI-YEDGVKRLDNDMDIIOLEAETIRKASQDDEARRI 365
DB 314 VNGPGALMTDIGFTGGAGAGVYVNAKRPDHKQSNEGMIDHIVLVERKKAETIOAKQ 373
QY 366 D 366
DB 374 D 374

Search completed: November 23, 2003, 17:15:28
Job time : 18 secs

AC P57374; 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
 GN ISPG OR GCPE OR B0287;
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 symbiotic bacterium).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Buchnera.
 OK NCBI_TaxID=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 Buchnera sp. APS.";
 RL Nature 407:81-86(2000).
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
 (ME-2,4CP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 (By similarity).
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway, sixth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB001118; BAB1297.1; -.
 DR HSP: P49058; IEP.
 DR HAMAP: MF_00159; -; 1.
 DR InterPro: IPR004588; ISPG.
 DR Pfam: PF04551; GCPE; 1.
 DR TIGRFAMs: TIGR00612; ispg GCPE; 1.
 KM Isoprene biosynthesis; Complete proteome.
 SQ SEQUENCE 368 AA; 40869 MW; 88C2F495954C1DE CRC64;
 Query Match 68.7%; Score 1282; DB 1; Length 368;
 Best Local Similarity 66.1%; Pred. No. 2.9e-81;
 Matches 242; Conservative 65; Mismatches .57; Indels 2; Gaps 2;
 QY 1 MHNQAF-IORRSTRIVGNVPIGDCAPFAVOSMTTRTTDEATVNOIKALERVAGADIV 59
 DB 1 MNNKCKIINRRKSDRIYGVKVAIGNNAPISVOSMTTRTTINSETINQILEQKVGVDIV 60
 QY 60 RVSVPTMDAAEAFKLIKQOVNPLVADHFDYRIALKAVEYGVDCIRIPNGINERIR 119
 DB 61 RSTIPMLKAESKEIKQTNVPLADHFDYRLAQAIRKYGDCIRIPNGINERIR 120
 QY 120 MVVDCARDKNIPIRIGVNGSLEKDOEKYGEPTPOALLSARHVDHLDRLNFDPKYS 179
 DB 121 EISYAKDENIPRIGVNGSLEKDLKKYKPTPALVESAMRHAEFYDALNFDPKYS 180
 QY 180 VKASDVFLAVESYRLAKIIDQPLHIGITEAGASGAVKSAIGLLSEIGIDTLRVS 239
 DB 181 VKASDVFLAVESYRLAKIIDQPLHIGITEAGASGAVKSAIGLLSEIGIDTLRVS 240
 QY 240 LAADPVEETIKVGDIKSLRISRGINFIACPTCSROEPDVIGTVNALSQRLEDITTPMD 299
 DB 241 LAADPVEETIKVGDIKSLRISRGINFIACPTCSROEPDVIGTVNALSQRLEDITTPMD 300
 QY 300 VSIIGCVNPGSALVSTLGTGNGKSGLYEDGVK-KDRLDNNMDIOLLEARIKAKSO 358
 DB 301 VSIIGCVNPGSALVSTLGTGNGKSGLYEDGVK-KDRLDNNMDIOLLEARIKAKSO 360
 QY 359 LDEARR 364
 DB 361 LNNKK 366

RESULT 14
 ID ISPG_WIGBR STANDARD; PRT; 366 AA.
 AC Q8DIY3;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
 GN ISPG OR GCPE OR WIGBR5730.
 OS Wigglesworthia glossinidia brevipalpis.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Wigglesworthia.
 OK NCBI_TaxID=36870;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=22297718; PubMed=12219091;
 RA Akman L., Yamashta A., Watanabe H., Oshima K., Shiba T., Hattori M.,
 RA Aksoy S.;
 RT "Genome sequence of the endocellular obligate symbiont of tsetse
 RT flies, Wigglesworthia glossinidia.";
 RL Nat. Genet. 32:402-407(2002).
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
 (ME-2,4CP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 (By similarity).
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway, sixth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB063522; BAC24719.1; -.
 DR HAMAP: MF_00159; -; 1.
 DR InterPro: IPR006705; GCPE.
 DR InterPro: IPR004588; ISPG.
 DR Pfam: PF04551; GCPE; 1.
 DR TIGRFAMs: TIGR00612; ispg GCPE; 1.
 KM Isoprene biosynthesis; Complete proteome.
 SQ SEQUENCE 366 AA; 40407 MW; 159B0169B8FC0DED CRC64;
 Query Match 65.7%; Score 1226.5; DB 1; Length 366;
 Best Local Similarity 65.9%; Pred. No. 1.9e-77;
 Matches 234; Conservative 62; Mismatches 58; Indels 1; Gaps 1;
 QY 7 IORRSTRIVGNVPIGDCAPFAVOSMTTRTTDEATVNOIKALERVAGADIVSVPTM 66
 DB 7 IIRRSKTIYINVPICGSPISVOSMTTCCTDINSTISQINKLOKACADIVRSIPTL 66
 QY 67 DAAEAFKLIKQOVNPLVADHFDYRIALKAVEYGVDCIRIPNGINERIRMVDCAR 126
 DB 67 EAAESFKTIKRVNVPPIVADHFDYRIALKAVEYGVDCIRIPNGINERIRIISVSTAK 126
 QY 127 DKNIPRIGVNGSLEKDOEKYGEPTPOALLSARHVDHLDRLNFDPKYSVKASDV 186
 DB 127 EKKLPIRIGVNGSLEKDIENKYGINNPKIPESAMRHVNILEKINFDWFKYSVSSDV 186
 QY 187 LAVESYRLAKIIDQPLHIGITEAGASGAVKSAIGLLSEIGIDTLRVS LAADPVE 246
 DB 187 TVQSYKVLAKIIDQPLHIGITEAGSMLHGSIKSIGLLSEIGIDTLRVS LAADPVE 246
 QY 247 EIKVGFDIKSLRISRGINFIACPTCSROEPDVIGTVNALSQRLEDITTPMDVSIIGCV 306
 DB 247 EYKVGFSILRSINIKRGINFIACPTCSROEPDVIGTVNALSQRLEDITTPMDVSIIGCV 306
 QY 307 VNGPGEALVSTLGTGNGKSGLYEDGVK-KDRLDNNMDIOLLEARIKAKSO 360
 DB 307 VNGLGEANRADIGISGRNKSILFENGGLRNNKINNEETIIDLEKYIRKVKILN 361

"whole-genome random sequencing and assembly of Haemophilus influenzae Rd."
 RT Science 269:496-512 (1995).
 RL
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
 CC (ME-2,4GPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 CC (By similarity).
 CC -1- PATHWAY: Nomenclature terpenoid biosynthesis pathway; sixth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, U32721; AAC22026.1; -.
 CC PIR, H64063; H64063.
 CC TIGR, H10368; -.
 CC HAMAP, MF_00159; -; 1.
 CC InterPro: IPR004588; ISPG.
 CC Pfam: PF04551; GCPE; 1.
 CC TIGRPFAM: TIGR00612; ispg_gcpe; 1.
 CC Isoprene biosynthesis; Complete proteome.
 KW
 SQ SEQUENCE 368 AA; 4016 MW; DC99EAL5A50ED01B CRC64;

Query Match 85.9%; Score 1603.5; DB 1; Length 368;
 Best Local Similarity 88.2%; Pred. No. 2.1e-103;
 Matches 320; Conservative 20; Mismatches 22; Indels 1; Gaps 1;

QY 4 QAPIORKSTRIVGNVPIGDAPIAVOSMTNRTTDEATVNOIKALERVGADIVRSV 63
 DB 5 OPTIKRRESTKIYGVNVPVIGDAPVAVQSMTRTTDVEATVQISLERVAGADIVRSV 64
 QY 64 PTMDAAEAKRLIKQOVNPLVADHFDYRIALKVAEYGVDCIRINPGNIGNEERIMVVD 123
 DB 65 PTMDAAEAKRQIKQOVNPLVADHFDYRIALKVAEYGVDCIRINPGNIGNEERIMVVD 124
 QY 124 CARDKPIRIGVNASLEKDOEKYGEPTPOLLESAMRHVDHRLNFDQKYSVKAS 183
 DB 125 CARDKPIRIGVNASLEKDOEKYGEPTPOLLESAMRHVDHRLNFDQKYSVKAS 184
 QY 184 DVFLLAVESYRLAKQIDQPLHIGITEAGAGSAGVSAIGLGLSEIGDITLRSVLAAD 243
 DB 185 DVFLLAVESYRLAKQIDQPLHIGITEAGAGSAGVSAIGLGLSEIGDITLRSVLAAD 244
 QY 244 PVEEIVGVFDILKSLRISRGINFACPTCSROEPDVGITVNALEQRLIEDITPMDVSI 303
 DB 245 PVEEIVGVFDILKSLRISRGINFACPTCSROEPDVGITVNALEQRLIEDITPMDVSI 304
 QY 304 GGVNPGALVSTLGVGTGNKKSGLYEDGVR-KDRLNNMDIMDQLEAIRAKASQLDEA 362
 DB 305 GGVNPGALVSDLGVTGNKKSGLYEDGVRKDFNEDIVNQLAKIRAKVARQDPK 364
 QY 363 RRI 365
 DB 365 NRI 367

RESULT 10
 ISPG_PASMU STANDARD; PRT; 367 AA.
 ID ISPG_PASMU
 AC P57987;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
 GN ISPG OR GCPE OR PM2010.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
 OC Pasteurellaceae; Pasteurella.
 NCBI_TaxID=747;

RN [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=pm70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida pm70."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
 CC (ME-2,4GPP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
 CC (By similarity).
 CC -1- PATHWAY: Nomenclature terpenoid biosynthesis pathway; sixth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL, AE006237; AA004094.1; -.
 CC HAMAP, MF_00159; -; 1.
 CC InterPro: IPR004588; ISPG.
 CC Pfam: PF04551; GCPE; 1.
 CC TIGRPFAM: TIGR00612; ispg_gcpe; 1.
 CC Isoprene biosynthesis; Complete proteome.
 KW
 SQ SEQUENCE 367 AA; 39969 MW; 1918C5702642AE8 CRC64;

Query Match 85.7%; Score 1599.5; DB 1; Length 367;
 Best Local Similarity 88.9%; Pred. No. 4e-103;
 Matches 320; Conservative 18; Mismatches 21; Indels 1; Gaps 1;

QY 7 IQRRKSTRIVGNVPIGDAPIAVOSMTNRTTDEATVNOIKALERVGADIVRSVPTM 66
 DB 7 IKRRSTKIYGVNVPVIGDAPVAVQSMTRTTDVEATVQIKALERVGADIVRSVPTM 66
 QY 67 DAAEAFKLIKQOVNPLVADHFDYRIALKVAEYGVDCIRINPGNIGNEERIMVVD 126
 DB 67 DAAEAFKLIKQOVNPLVADHFDYRIALKVAEYGVDCIRINPGNIGNEERIMVVD 126
 QY 127 DKNIPRIGVNASLEKDOEKYGEPTPOLLESAMRHVDHRLNFDQKYSVKASDVF 186
 DB 127 DKNIPRIGVNASLEKDOEKYGEPTPOLLESAMRHVDHRLNFDQKYSVKASDVF 186
 QY 187 LAVESYRLAKQIDQPLHIGITEAGAGSAGVSAIGLGLSEIGDITLRSVLAADPVE 246
 DB 187 LAVESYRLAKQIDQPLHIGITEAGAGSAGVSAIGLGLSEIGDITLRSVLAADPVE 246
 QY 247 EIKVGFILKSLRISRGINFACPTCSROEPDVGITVNALEQRLIEDITPMDVSI 306
 DB 247 EIKVGFILKSLRISRGINFACPTCSROEPDVGITVNALEQRLIEDITPMDVSI 306
 QY 307 VNGPGEALVSTLGVGTGNKKSGLYEDGVR-KDRLNNMDIMDQLEAIRAKASQLDEARRI 365
 DB 307 VNGPGEALVSDLGVTGNKKSGLYEDGVRKDFNEDIVNQLAKIRAKVARQDPK 366

RESULT 11
 ISPG_PSEAE STANDARD; PRT; 371 AA.
 ID ISPG_PSEAE
 AC O9HXJ4;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
 GN ISPG OR PA3803.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 NCBI_TaxID=287;
 RN [1]
 CC SEQUENCE FROM N.A.

KW Isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 375 AA; 40797 MW; 860971872B1C0536 CRC64;

Query Match 87.0%; Score 1622.5; DB 1; Length 375;
Best Local Similarity 86.8%; Pred. No. 1.1e-104;
Matches 323; Conservative 19; Mismatches 29; Indels 1; Gaps 1;

QY 1 MHNQAPIQRKSTRIVGVNPIGDGAPIAVQSMNTNRTTDTVEATVQIKALERVGADIVR 60
DB 1 MNGSPTIRKSRIRIVGVNPIGDGAPIAVQSMNTNRTTDTVEATVQIKALERVGADIVR 60
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAAYGVDCLRINPQNGNERIRM 120
DB 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAAYGVDCLRINPQNGNERIRM 120
QY 121 VVDCARDKXNIPRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDLRLNPDQKVS 180
DB 121 VVACARDYNIPIRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDLRLNPDQKVS 180
QY 181 KASDVFLAVESYRLAKQIDQPLHLGITEAGARGSAVSAIGLLSEIGDTRLVSL 240
DB 181 KASDVFLAVESYRLAKQIDQPLHLGITEAGARGSAVSAIGLLSEIGDTRLVSL 240
QY 241 AADPVEEIKVGFILKSLRIRSGINFIACPTCSROEPFVIGTVNALQRLLEDITPM 300
DB 241 AADPVEEIKVGFILKSLRIRSGINFIACPTCSROEPFVIGTVNALQRLLEDITPM 300
QY 301 SIIGCVNPGELAVSTLCTGANKSKSGLYEDEVK-KDRLDNDMDLDLEARRAASQL 359
DB 301 SIIGCVNPGELAVSTLCTGANKSKSGLYEDEVK-KDRLDNDMDLDLEARRAASQL 359
QY 360 DEARRIDVOQVE 371
DB 361 DANNRIVINQLD 372

RESULT 8

ISPG_VIBCH STANDARD; PRT; 376 AA.

AC 09KTYL, 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ISPG OR VC0759.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tetrelin H., Richardson D.,
RA Ermoiaeva M.D., Vamathevan J., Bass S., Qin H., Dracot I., Sellers P.,
RA McDonald L., Utechtack T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RA "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae."
RT Nature 406:477-483(2000).
RL - FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4CpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (BY similarity).
CC - PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
CC - SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL, AF004161; AAF93924.1; --
DR PIR; F82283; F82283.
DR TIGR; VC0759; --
DR HAMAB; MF_00159; --
DR InterPro; IPR004588; ISPG.
DR Pfam; PF04551; GCPE; 1.
DR TIGRPFAMs; TIGR00612; ISPG_gcpe; 1.
KW Isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 376 AA; 40862 MW; 34E4144CB87ADAAA CRC64;

Query Match 86.9%; Score 1621.5; DB 1; Length 376;
Best Local Similarity 85.3%; Pred. No. 1.3e-104;
Matches 318; Conservative 33; Mismatches 21; Indels 1; Gaps 1;

QY 1 MHNQAPIQRKSTRIVGVNPIGDGAPIAVQSMNTNRTTDTVEATVQIKALERVGADIVR 60
DB 3 MGHESPTRKSRIRIVGVNPIGDGAPIAVQSMNTNRTTDTVEATVQIKALERVGADIVR 62
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAAYGVDCLRINPQNGNERIRM 120
DB 63 VSVPTMDAAEAFKLIKQOVNPLVADIHFDYRIALKAAYGVDCLRINPQNGNERIRM 122
QY 121 VVDCARDKXNIPRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDLRLNPDQKVS 180
DB 123 VVDCARDKXNIPRIGVNGSLKEDLOEKYGEPTPOLLESAMRHVDLRLNPDQKVS 182
QY 181 KASDVFLAVESYRLAKQIDQPLHLGITEAGARGSAVSAIGLLSEIGDTRLVSL 240
DB 183 KASDVFLAVESYRLAKQIDQPLHLGITEAGARGSAVSAIGLLSEIGDTRLVSL 242
QY 241 AADPVEEIKVGFILKSLRIRSGINFIACPTCSROEPFVIGTVNALQRLLEDITPM 300
DB 243 AADPVEEIKVGFILKSLRIRSGINFIACPTCSROEPFVIGTVNALQRLLEDITPM 302
QY 301 SIIGCVNPGELAVSTLCTGANKSKSGLYEDEVK-KDRLDNDMDLDLEARRAASQL 359
DB 303 SIIGCVNPGELAVSTLCTGANKSKSGLYEDEVK-KDRLDNDMDLDLEARRAASQL 362
QY 360 DEARRIDVOQVE 372
DB 363 DEKRRIDKIHVQD 375

RESULT 9

ISPG_HAEIN STANDARD; PRT; 368 AA.

AC P44667, 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ISPG OR GCPE OR H10368.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=127;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utechtack T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;


```

QY 301 SIIGCVNPGGALVSTLTGVTGKNGKSGLYEDGVR-KDRLDNNMDIDLEARIKASQUL 359
DB 301 SIIGCVNPGGAEVSHLGLAGSNKKSAYEDQKQKERFEDNDLYNQLEAKIRAKAAM 360
QY 360 DEARRIDVQ 368
DB 361 DESNRIDIK 369

RESULT 6
ISPG_PROST STANDARD; PRT; 365 AA.
ID ISPG_PROST
AC P72241;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ISPG OR AAC.
OS Providencia stuartii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Providencia.
OX NCBI_TaxId=588;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=PR50;
RX MEDLINE=97234638; PubMed=9079912;
RA Rather P.N., Solinsky K.A., Paradise M.R., Parojic M.M.;
RT "aacC, an essential gene involved in density-dependent regulation of
RT the 2'-N-acetyltransferase in Providencia stuartii."
RL J. Bacteriol. 179:2267-2273(1997).
CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4cnp) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (By similarity). Involved in density-dependent regulation of 2'-N-
CC acetyltransferase.
CC -1- PATHWAY: Nomenclature terpenoid biosynthesis pathway; sixth step.
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U67933; AAB51469.1; -
CC HAMAP; MF_00159; -; 1.
CC InterPro; IPR004588; ISPG.
CC Pfam; PF04551; GCPE; 1.
CC TIGRfam; TIGR00612; ispg_gcpe; 1.
CC Isoptrene biosynthesis.
CC SEQUENCE 365 AA; 39840 MW; 4523147980480ECC CRC64;

Query Match 87.0%; Score 1622.5; DB 1; Length 365;
Best Local Similarity 88.4%; Pred. No. 1e-104;
Matches 320; Conservative 24; Mismatches 17; Indels 1; Gaps 1;

```

```

QY 241 AADPVEIKVGFIDILKSIRIRSGINFIACPTCSRQEPDVGTVNALEQRLEDITPMV 300
DB 241 AADPVEIVGVGDILKSIRIRSGINFIACPTCSRQEPDVGTVNALEQRLEDITPMV 300
QY 301 SIIGCVNPGGALVSTLTGVTGKNGKSGLYEDGVR-KDRLDNNMDIDLEARIKASQUL 359
DB 301 SIIGCVNPGGAEVSHLGLAGSNKKSAYEDQKQKERFEDNDIINDLEAKIRAKAAM 360
QY 360 DE 361
DB 361 DE 362

RESULT 7
ISPG_YERPE STANDARD; PRT; 375 AA.
ID ISPG_YERPE
AC P58672;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN ISPG OR GCPE OR IPO2879 OR Y1353.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxId=632;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parhill J., Wren B.W., Thomson N.R., Churcher C., Mungall K.L.,
RA Prentice M.B., Sebatia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks R.M., Davis P., Dougan G.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Meule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague."
RL Nature 413:523-527(2001).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=KIM5 / Biovar Mediaevalis;
RX MEDLINE=2213763; PubMed=12142430;
RA Deng M., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL J. Bacteriol. 184:4601-4611(2002).
CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4cnp) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (By similarity).
CC -1- PATHWAY: Nomenclature terpenoid biosynthesis pathway; sixth step.
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ41454; CAC92130.1; -
CC EMBL; AB013738; AAM84926.1; -
CC PIR; AG0350; AG0350.
CC HAMAP; MF_00159; -; 1.
CC InterPro; IPR006705; GCPE.
CC InterPro; IPR004588; ISPG.
CC Pfam; PF04551; GCPE; 1.
CC TIGRfam; TIGR00612; ispg_gcpe; 1.

```

```

Db      301 SIICGVNNGPEALVSTLGTVGANKKSGLYEDVRKORLNDMDMIAQIESIRAKVSGLD 360
Oy      361 EARRIDVQVQEK 372
        |||||
Db      361 EARRIDVQVQEK 372

RESULT 4
ISPG_VIBPA STANDARD, PRT, 372 AA.
ID  ISPG_VIBPA
AC  087S16;
DT  15-SEP-2003 (Rel. 42, Created)
DT  15-SEP-2003 (Rel. 42, Last sequence update)
DT  15-SEP-2003 (Rel. 42, Last annotation update)
DE  1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
DE  ISPG OR VP0608.
OS  Vibrio parahaemolyticus.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC  Vibrionaceae; Vibrio.
OX  NCBI_TaxID=670;
RX  MEDLINE=22508454; Pubmed=12620739;
RX  Mafino K., Ogihara K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA  Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA  Yabuaga T., Honda T., Shinagawa H., Hattori M., Iida T.;
RT  "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT  distinct from that of V. cholerae.";
RL  Lancet 361:743-749(2003).
CC  -i- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC  (ME-2, 4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC  (By similarity).
CC  -i- PATHWAY: Nomenclonate terpenoid biosynthesis pathway; sixth step.
CC  -i- SIMILARITY: BELONGS TO THE ISPG FAMILY.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
CC  EMBL, AP005075; BAC56871.1; -.
DR  HAMAP, MF_00159; -.
KW  Isoprene biosynthesis; Complete proteome.
SQ  SEQUENCE 372 AA; 40688 MW; EAAFAAALAT64DDB8 CRC64;

Query Match: 88.2%; Score 1645.5; DB 1; Length 372;
Best Local Similarity 87.5%; Pred. No. 2.8e-106;
Matches 323; Conservative 27; Mismatches 18; Indels 1; Gaps 1.

Oy      1 MHNQAPIORRKRSTRIVYGVNPIGDGAPLAVGSMTRTRTDEATVNOIKALERVGADIVR 60
        -:::-|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      1 MGHSEPIIRKSTRIVYGVNPIGDGAPLAVGSMTRTRTDEATVNOIKALERVGADIVR 60
Oy      61 VSVPTMDAAEAKLIKQOVNVPVLVDIHEDYIALKVAEYGDICRINPGNIGNERIRM 120
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      61 VSVPTMEAAEAKLIKQOVSVPLVADIHEDYIALKVAEYGVDCRINPGNIGNERIRS 120
Oy      121 VVDCARDKNIPIRIGVNGSLEKDIQEKXGEPTPOLLESARHVDHLDRINFPDFKYSV 180
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      121 VVDCARDKNIPIRIGVNGSLEKDIQMKYGEPTPEALVESAKRHVDHLDRINFPDFKYSV 180
Oy      181 KASDVFLAVESYRLAKOIDPLHIGITBAGARGASGAVKSAIGLILSEIGIDTLRVSL 240
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      181 KASDVFLAVDSYRLAKRIDPLHIGITBAGARGAGAVKSAVGLGWLISEGIDTLRISL 240
Oy      241 AADPVEETIVGDDILKSLRIBSRGINFICQPCSCQEPFIVGTVALREQRLEDITTPMDV 300
        |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db      241 AADPVEETIVGDDILKSLRIBSRGINFICQPCSCQEPFIVGTVALREQRLEDITTPMDV 300

```

QY	301	SIICVANGGEALVSTLGTGAKKSGKGLVEDGVR-KDRLDNNMIMQLEKRIAPKASQI	359
DB	301	SIICVANGGEALVSHLGLAGSKKSAFVEDGRQKERFPDNDLVNQLEKIRAKASMM	360
QY	360	DEARRIDVQ 368	
DB	361	DSERIEIK 369	
RESULT 5			
ISPG_VIBVU	ISPG_VIBVU	STANDARD;	PRT; 372 AA.
AC	Q8DEZ8:		
DT	15-SEP-2003	(Rel. 42, Created)	
DT	15-SEP-2003	(Rel. 42, Last sequence update)	
DT	15-SEP-2003	(Rel. 42, Last annotation update)	
DE	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.		
GN	ISPG OR V10427.		
OS	Vibrio vulnificus.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;		
OC	Vibrionaceae; Vibrrio.		
OX	NCBI_TaxID=672;		
RA	SEQUENCE FROM N.A.		
RC	STRAIN-CMCP6;		
RA	Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,		
RT	Choy H.E.;		
RL	"Complete genome sequence of Vibrio vulnificus CMCP6.";		
CC	Submitted (DEC-2002) to the EMBL/Genbank/DBJ databases.		
CC	-1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate		
CC	(ME-2,4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate		
CC	(by similarity).		
CC	-1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.		
CC	-1- SIMILARITY: BELONGS TO THE ISPG FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	at the European Bioinformatics Institute. There are no restrictions on		
CC	its use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ib-sib.ch).		
CC	or send an email to license@ib-sib.ch).		
CC	-----		
DR	EMBL; AE016798; AAC08950.1; -		
DR	HMAP; MF 00159; -; 1.		
DR	Pfam; PF04551; GcPE; 1.		
DR	TIGRFAMs; TIGR00612; ispg_gcpe; 1.		
KM	Sequence biosynthesis; Complete proteome.		
SQ	SEQUENCE 372 AA; 40573 MW; 6C73C2BB8619B0285 CRC64;		
Query Match 87.4%; Score 1631.5; DB 1; Length 372;			
Best Local Similarity 86.7%; Pred. No. 2.6e-105;			
Matches 320; Conservative 29; Mismatches 19; Indels 1; Gaps 1.			
QY	1	MENQAPICRRKSTRIVYGVNVPICGAPAVOSMTNTRTTDVEATYNQIKALERVADIVR	60
DB	1	MHNESPIIRRSSTRIVYGVDPICGAPAVOSMTNTRTTDVAATVAQIKALENVGADIVR	60
QY	61	VSVPFMDAAEAFKLIKQOVNPIVADIHFDRIAKVAEYGVDCIRINPGINNEERIRM	120
DB	61	VSVPFMDAAEAFKLIKQOVNPIVADIHFDRIAKVAEYGVDCIRINPGINNEERIRS	120
QY	121	VVDCARDKNIPIRISVNGSLEKDOEQEYGEPTPALLESAMRHVDLRINFDFQKVS	180
DB	121	VVDCARDKNIPIRISVNGSLEKDOEQEYGEPTPALLESAMRHVDLIDRLNFDQKVS	180
QY	161	KASDVFAVESYRIIAKOIDOPHLGITBAGARGSAVKSAIGLLISEGIGPTLRVSL	240
DB	161	KASDVFAVDSYRIIAKRIDOPHLGITBAGARGSAVKSAVGLMLAEGIGDTLRISL	240
QY	241	AADPVEELKVGSDILKSLRISRGINFIACPTCSQOEFDVGTVALBQRLEDITTPMDV	300
DB	241	AANPVEELKVGSDILKSLRISRGINFIACPSQOEFDVISTVALBQRLEDITTPMDV	300

```

RT      LTY.: " ;
RL      Nature 413:852-856(2001) .
CC      -I- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC      (ME-2,4cpe) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC      (By samilarity) .
CC      -I- PATHWAY: Nomevalonate terpenoid biosynthesis pathway; sixth step.
CC      -I- SIMILARITY: BELONGS TO THE ISPG FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch) .
CC
CC      EMBL; AE008814; AAL21417.1; -.
DR      StyGene; SG?/?/?/?; ispg.
DR      HAMAP; MF_00159; -; 1.
DR      InterPro; IPR0046705; ispg.
DR      InterPro; IPR004588; ispg.
DR      Pfam; PF04551; Gcpe; 1.
DR      TIGRFAMs; TIGR00612; ispg_gcpe; 1.
DR      KEGG; ispgene_biosynthesis; Complete proteome.
SQ      SEQUENCE   372 AA; 40625 MW; 0486FD3BCECA2A9F CRC64;

Query Match          98.6%; Score 1840; DB 1; Length 372;
Best Local Similarity 98.7%; Pred. No. 1.le-119;
Matches 367; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      1 MHNAPRIORRKSTIIYGVNPVIGGAPIAVGOSMTNTRTDVEATVNOIKALERGADIVR 60
DB      |||
QY      1 MHNAPRIORRKSTRIYGVNPVIGGAPIAVGOSMTNTRTDVEATVNOIKALERGADIVR 60
DB      |||
QY      61 VSVPTMDAAEFKLIKQOVNPLVADIHFDPRIALKVAEYVDCGRINPVGIGNEERIRM 120
DB      |||
QY      61 VSVPTMDAAEFKLIKQOVNPLVADIHFDPRIALKVAEYVDCGRINPVGIGNEERIRM 120
DB      |||
QY      121 VVDCARDKNIPIRIGVNAAGLEKDLQEKYGEPTQALLLESAMRHVDLRINFDOFRYSV 180
DB      |||
QY      121 VVDCARDKNIPIRIGVNAAGLEKDLQEKYGEPTQALLLESAMRHVDLRINFDOFRYSV 180
DB      |||
QY      181 KASVPFLAVESRYLLAQIDOPHLGITTEAGAGSAGVAKSAIGGLLSBGITGTLRVSL 240
DB      |||
QY      181 KASVPFLAVESRYLLAQIDOPHLGITTEAGAGSAGVAKSAIGGLLSBGITGTLRVSL 240
DB      |||
QY      241 AADVEEIKYGFDLKSLRISRCGINFIACPTGSRGFEDVGTGNALBOQLIEDIITPMDV 300
DB      |||
QY      241 AADVEEIKYGFDLKSLRIRAKGINFIACPTGSRGFEDVGTGNALBOQLIEDIITPMDV 300
DB      |||
QY      301 SIICGVNNGPEALVSTLGVTGGNKSGLYEDGVKRDLDDNDMVIDOLEAIRAKASOLD 360
DB      |||
QY      301 SIICGVNNGPEALVSTLGVTGGNKSGLYEDGVKRDLDDNDMVIDOLEAIRAKASOLD 360
DB      |||
QY      361 EARRIDVQYOEK 372
DB      |||||
QY      361 EARRIDVLOVER 372
DB      |||||

RESULT 3
ISPG_SALT1
ID      ISPG_SALT1 STANDARD; PRT; 372 AA.
AC      P58670;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      15-SEP-2003 (Rel. 42, Last annotation update)
DE      1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
GN      ispg OR GCPE OR STY2768 OR TO333.
OS      Salmonella typhi.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Salmonella.
NCBI_TaxId=601;
LN      [1]

```

```

RP SEQUENCE FROM N.A.
RC
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor J.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hen T.T., Holtroyd S., Jagels K.,
RA Krogsh A., Larsen T.S., Leachter S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RT Nature 413:848-852(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Ty2 / ATCC 700931;
RX MEDLINE=22531367; PubMed=12644504;
RA Deng W., Liou S.-R., Plunkett G., III, Mayhew G.F., Rose D.J.,
RA Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R.;
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2
RT and CT18."
RL J. Bacteriol. 185:2330-2337(2003).
CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
CC (ME-2,4cpP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate
CC (by similarity).
CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, AL627275; CAD02726.1; -
DR EMBL, AE016835; AAO68056.1; -
DR HAMAP, MF_00159; -; 1.
DR InterPro: IPR006705; Gcpe.
DR InterPro: IPR004588; Ispg.
DR Pfam, PF04551; Gcpe; 1.
DR TIGRFAMs, TIGR00612; ispg_gcpe; 1.
DR KW Isoprene biosynthesis; Complete proteome.
SQ SEQUENCE 372 AA; 40626 MW; 3E3DB13BAA7916 CRC64;

Query Match 98.1%; Score 1831; DB 1; Length 372;
Best Local Similarity 98.1%; Pred. No. 4.8e-119;
Matches 365; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MHNQAPIORRKSRIYGVNVPVIGDGAPIAQSMTNTFTTDEATVNOIKALERVGADIVR 60
DB 1 MHNQAPIORRKSRIYGVNVPVIGDGAPIAQSMTNTFTTDEATVNOIKALERVGADIVR 60
QY 61 VSVPTMDAAEFLLKIQOVNVPLVADIHEDYRIALKYAEYGVDCRLINPGNIGESIRIM 120
DB 61 VSVPTMDAAEFLLKIQOVNVPLVADIHEDYRIALKYAEYGVDCRLINPGNIGESIRIM 120
QY 121 VVDCARDKNPIRITGVNAGSLKEDLDKQKEPFPQALLESAMRHVDLRLNPFQFKSV 180
DB 121 VVDCARDKNPIRITGVNAGSLKEDLDKQKEPFPQALLESAMRHVDLRLNPFQFKSV 180
QY 181 KASDVFLLAVESYRLAKQIDQPLHLGITTEGARGSGVAKAIGLLLSGIGDTLRVSL 240
DB 181 KASDVFLLAVESYRLAKQIDQPLHLGITTEGARGSGVAKAIGLLLSGIGDTLRVSL 240
QY 241 AADPVEEIKVGFIDILKSLRIRSGINFIACPTCSRQEFVIGTVNALEQLLEDIITPM 300
DB 241 AADPVEEIKVGFIDILKSLRIRSGINFIACPTCSRQEFVIGTVNALEQLLEDIITPM 300
QY 301 SIIGCVNNGELAVSTLTGTGNGKKSGLYEDVQRDRLDNNMDIDLEKIRIRAKSOLD 360
DB 301 SIIGCVNNGELAVSTLTGTGNGKKSGLYEDVQRDRLDNNMDIDLEKIRIRAKSOLD 360

```

RP SEQUENCE FROM N.A.
 RC STRAIN-06.H / CFT073 / ATCC 700928;
 RA MEDLINE=22388234; PubMed=12471157;
 RA Welch R.A., Burland V., Plunkett G., III, Redford P., Roesch P.,
 RA Rasko D., Buckles E.L., Liu S.-R., Boutin J., Hackett J., Stroud D.,
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
 RT "Extensive mosaic structure revealed by the complete genome sequence
 of uropathogenic *Escherichia coli*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN-0157.H7 / BDL933 / ATCC 700927;
 RA MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimallanta E.T., Potamalis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC STRAIN-0157.H7 / RIMD 0509952;
 RA MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa G., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 RN [8]
 RP PRELIMINARY SEQUENCE OF 343-372 FROM N.A.
 RA MEDLINE=85261421; PubMed=2991272;
 RA Freedman R., Gibson B., Donovan D., Blemann K., Eisenbeis S.J.,
 RA Parker J., Schimmel P.;
 RT "Primary structure of histidine-tRNA synthetase and characterization
 of his^c transcripts.";
 RL J. Biol. Chem. 260:10063-10068(1985).
 RN [9]
 RP PATHWAY.
 RC STRAIN-K12 / MC4100;
 RA MEDLINE=21099853; PubMed=11163766;
 RA Campos N., Rodriguez-Concepcion M., Seemann M., Boronat A.;
 RT "Identification of gcpe as a novel gene of the 2-C-methyl-D-erythritol
 4-phosphate pathway for isoprenoid biosynthesis in *Escherichia coli*.";
 RL FEBS Lett. 488:170-173(2001).
 RN [10]
 RP PATHWAY.
 RA MEDLINE=21172855; PubMed=11274098;
 RA Altincicek B., Kollas A.-K., Sanderbrand S., Wiesner J., Hintz M.,
 RA Beck E., Jomaa H.;
 RT "Gcpe is involved in the 2-C-methyl-D-erythritol 4-phosphate pathway
 of isoprenoid biosynthesis in *Escherichia coli*.";
 RL J. Bacteriol. 183:2411-2416(2001).
 CC -1- FUNCTION: Converts 2C-methyl-D-erythritol 2,4-cyclodiphosphate
 (ME-2,4CP) into 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate.
 CC -1- PATHWAY: Nonmevalonate terpenoid biosynthesis pathway; sixth step.
 CC -1- SIMILARITY: BELONGS TO THE ISPG FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; X6451; CAA45783.1;
 DR EMBL; AY03515; AAK53460.1;
 DR EMBL; AE000338; AAC75568.1;

DR EMBL; D90880; BAA16402.1;
 DR EMBL; D90881; BAA20919.1;
 DR EMBL; AE016764; FAN81487.1;
 DR EMBL; AE005481; AAG57625.1;
 DR EMBL; AP002561; BAB36800.1;
 DR EMBL; M1843; NOT_ANNOTATED_CDS.
 DR PIR; A91051; A91051.
 DR PIR; E85895; E85895.
 DR PIR; S23058; S23058.
 DR EcoGene; EGI0370; ispG.
 DR HAMAP; MF_00159; -; 1.
 DR InterPro; IPR004588; Ispg.
 DR Pfam; PF04551; GCPE; 1.
 DR TIGRfams; TIGR00612; ispg_gcpe; 1.
 DR Isoprene biosynthesis; Complete proteome.
 SW SEQUENCE 372 AA; 40683 MW; 3666F83D5CD4F9F CRC64;
 Query Match 100.0%; Score 1866; DB 1; Length 372;
 Best local Similarity 100.0%; Pred. No. 1.9e-121;
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MHNQAPIQRKSTRIYGVNPIGDGAPIAVQSMNTTTRTDEATVNOIALERVGADIYR 60
 DB 1 MHNQAPIQRKSTRIYGVNPIGDGAPIAVQSMNTTTRTDEATVNOIALERVGADIYR 60
 QY 61 VSPPTMDAAEARKLKQVNVPLVADHPDYRIALKVAYGVDCIRINQNGNERIRM 120
 DB 61 VSPPTMDAAEARKLKQVNVPLVADHPDYRIALKVAYGVDCIRINQNGNERIRM 120
 QY 121 VVDCARDKNIPIRIGVNASLEKDLQEKGEPTPOLLESARHVDLRLNFDQPKVSV 180
 DB 121 VVDCARDKNIPIRIGVNASLEKDLQEKGEPTPOLLESARHVDLRLNFDQPKVSV 180
 QY 181 KASDVFLAVESYRLAKQIDQPLHLGITBAGGARSQAVSAIGLLSEGGDTLRVSL 240
 DB 181 KASDVFLAVESYRLAKQIDQPLHLGITBAGGARSQAVSAIGLLSEGGDTLRVSL 240
 QY 241 AADPVEIKVGRFDILKSLRIRSRGINFACPTCSQGEFVIGTVNLQRLEDITPMQV 300
 DB 241 AADPVEIKVGRFDILKSLRIRSRGINFACPTCSQGEFVIGTVNLQRLEDITPMQV 300
 QY 301 SIICGVNPGGALVSTLGVGTGANKKSGLYEGVGRKDRLDNNDMDQLBARIRAKSQLD 360
 DB 301 SIICGVNPGGALVSTLGVGTGANKKSGLYEGVGRKDRLDNNDMDQLBARIRAKSQLD 360
 QY 361 EARRIDVQVEK 372
 DB 361 EARRIDVQVEK 372
 RESULT 2
 ISPG_SALTY
 ID ISPG_SALTY STANDARD; PRT; 372 AA.
 AC P58671;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase.
 GN ISPG OR GCPE OR STM2523.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 ON NCBI_TaxId=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2 / SGSC1412 / ATCC 700720;
 RA MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 15:55:37 ; Search time 17 Seconds

(without alignments)
1029.055 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866
Sequence: 1 MHNQAPIQRKSTRIVGVN.....RAKASQLEBARIDVQVEK 372

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	1866	100.0	372 1	ISPG_ECOLI
2	1840	98.6	372 1	ISPG_SALTY
3	1831	98.1	372 1	ISPG_SALTY
4	1645.5	88.2	372 1	ISPG_VIBPA
5	1631.5	87.4	372 1	ISPG_VIBVU
6	1622.5	87.0	372 1	ISPG_PROST
7	1622.5	87.0	372 1	ISPG_YERPE
8	1621.5	86.9	376 1	ISPG_VIBCH
9	1603.5	85.9	368 1	ISPG_HABIN
10	1599.5	85.7	371 1	ISPG_PSMU
11	1372.5	73.6	371 1	ISPG_PSEAE
12	1285.5	68.9	367 1	ISPG_BUCAP
13	1282	68.7	368 1	ISPG_BUCAI
14	1226.5	65.7	366 1	ISPG_WIGBR
15	1042.5	55.9	383 1	ISPG_CAVCR
16	906	48.6	349 1	ISPG_CLOPE
17	902.5	48.4	355 1	ISPG_THETN
18	887	47.5	369 1	ISPG_CLOAB
19	882.5	47.3	367 1	ISPG_BACHD
20	869.5	46.6	354 1	ISPG_FUSNO
21	868	46.5	368 1	ISPG_LISMO
22	859.5	46.1	377 1	ISPG_BACSU
23	838	44.4	357 1	ISPG_AQUAE
24	815.5	43.7	384 1	ISPG_STRCO
25	815	43.7	385 1	ISPG_STRCO
26	794	42.6	359 1	ISPG_HELPU
27	790	42.3	359 1	ISPG_HELPU
28	761	40.8	378 1	ISPG_CORGL
29	751.5	40.3	387 1	ISPG_WYCTU
30	712.5	38.2	392 1	ISPG_MYCLE
31	706	37.0	344 1	ISPG_THEMA
32	690	37.0	357 1	ISPG_CAMJE
33	618.5	33.1	403 1	ISPG_SYNY3

34	613.5	32.9	408 1	ISPG_ANASP	P58666 anabaena sp
35	588	31.5	404 1	ISPG_TREPA	O83460 treponema p
36	543.5	29.1	429 1	ISPG_DEIRA	O9rx99 deionococcus
37	527	28.2	421 1	ISPG_NEIMA	O9j14 neisseria m
38	527	28.2	421 1	ISPG_NEIMA	O9j240 neisseria m
39	512	27.4	417 1	ISPG_RHILU	O98t90 rhizobium l
40	509	27.3	420 1	ISPG_BRUME	O8y117 bruceella me
41	504.5	27.0	601 1	ISPG_CHLNU	O9pkx3 chlamydia m
42	493	26.4	417 1	ISPG_RHIME	O92119 rhizobium m
43	493	26.4	431 1	ISPG_RALSO	P58669 ralestonia s
44	492	26.4	416 1	ISPG_AGRTS	P58665 agrobacteri
45	487	26.1	602 1	ISPG_CHLUTR	O84060 chlamydia t

ALIGNMENTS

RESULT 1	ISPG_ECOLI	STANDARD;	PRT;	372 AA.
ID	P27433; P76984; P76985;			
AC	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DE	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase (Gcpe protein)			
GN	(Protein E).			
OS	ISPG OR GCPE OR B2515 OR C3037 OR Z3778 OR ECS3377.			
OS	Escherichia coli,			
OS	Escherichia coli, 06, and			
OS	Escherichia coli O157:H7.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562, 217992, 83334;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RX	MEDLINE=92394488; PubMed=152167;			
RA	Baker J., Franklin D.B., Parker J.;			
RT	"Sequence and characterization of the gcpe gene of Escherichia coli.;"			
RL	FEMS Microbiol. Lett. 73:1175-180(1992).			
RN	[2]			
RP	SEQUENCE FROM N.A., AND CHARACTERIZATION.			
RX	MEDLINE=97426617; PubMed=9278503;			
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,			
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,			
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,			
RA	Mau B., Shao Y.;			
RT	"The complete genome sequence of Escherichia coli K-12.;"			
RL	Science 277:1453-1474(1997).			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=K12;			
RX	MEDLINE=97349980; PubMed=9205837;			
RA	Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,			
RA	Itoh T., Kimura S., Kitegawa M., Makino K., Miki T., Mitsuhashi N.,			
RA	Oshibuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,			
RA	Mizushima T., Oyama S., Saito N., Sampei G., Satoch Y., Sivasubram S.,			
RA	Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,			
RA	Yamagata S., Horuchi T.;			
RT	"Construction of a contiguous 874-kb sequence of the Escherichia coli			
RT	K-12 genome corresponding to 50.0-68.8 min on the linkage map and			
RT	analysis of its sequence features.;"			
RL	DNA Res. 4:91-113(1997).			
RN	[5]			

THIS PAGE BLANK (USPTO)

QY 1 MHQAIPQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTTDVEATVNOIKALERVADIYR 60
DB 1 MHQAIPQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTTDVEATVNOIKALERVADIYR 60
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADHFDYRIALKAEGVDCRLNPNIGNEERRIM 120
DB 61 VSVPTMDAAEAFKLIKQOVNPLVADHFDYRIALKAEGVDCRLNPNIGNEERRIM 120
QY 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDQFKVSV 180
DB 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDQFKVSV 180
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSGAVSAIGLGLLSEIGDITLRVSL 240
DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSGAVSAIGLGLLSEIGDITLRVSL 240
QY 241 AADPVEIKVGFPIKSLRIRSGINPIACPTCSROEPDVTGTVNALBQRLIEDIITPMV 300
DB 241 AADPVEIKVGFPIKSLRIRSGINPIACPTCSROEPDVTGTVNALBQRLIEDIITPMV 300
QY 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNDMDIQLBARIKASQD 360
DB 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNDMDIQLBARIKASQD 360
QY 361 EARRIDVOQVEK 372
DB 361 EARRIDVOQVEK 372

RESULT 5
AG0350
probable acetyltransferase aarc [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #ext_change 09-Nov-2001
C:Accession: AG0350
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Tarata, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrett,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0350
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-375 <KUR>
A:Cross-references: GB:AL590842; PIDN:CA92130.1; PID:g15980846; GSPDB:GN00175
C:Genetics:
A:Gene: aarc
C:Superfamily: gcpb protein

Query Match 87.0%; Score 1622.5; DB 2; Length 375;
Best Local Similarity 86.8%; Pred. No. 9.1e-105;
Matches 333; Conservative 19; Mismatches 29; Indels 1; Gaps 1;

QY 1 MHQAIPQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTTDVEATVNOIKALERVADIYR 60
DB 1 MHQAIPQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTTDVEATVNOIKALERVADIYR 60
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADHFDYRIALKAEGVDCRLNPNIGNEERRIM 120
DB 61 VSVPTMDAAEAFKLIKQOVNPLVADHFDYRIALKAEGVDCRLNPNIGNEERRIM 120
QY 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDQFKVSV 180
DB 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDQFKVSV 180
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSGAVSAIGLGLLSEIGDITLRVSL 240
DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSGAVSAIGLGLLSEIGDITLRVSL 240
QY 241 AADPVEIKVGFPIKSLRIRSGINPIACPTCSROEPDVTGTVNALBQRLIEDIITPMV 300
DB 241 AADPVEIKVGFPIKSLRIRSGINPIACPTCSROEPDVTGTVNALBQRLIEDIITPMV 300

QY 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNDMDIQLBARIKASQD 360
DB 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNDMDIQLBARIKASQD 360
QY 360 DEARRIDVOQVE 371
DB 361 DANNRIVINQD 372

RESULT 6
F82283

gcpb protein VC0759 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #ext_change 02-Feb-2001
C:Accession: F82283
R:Reiherberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AF0035; MUID:20406833; PMID:10952301
A:Accession: F82283
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <HEI>
A:Cross-references: GB:AE004161; GB:AE003852; NID:g9655200; PIDN:AAF93924.1; GSPDB:GN001;
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0759
A:Map position: 1
C:Superfamily: gcpb protein

Query Match 86.9%; Score 1621.5; DB 2; Length 376;
Best Local Similarity 85.3%; Pred. No. 1.1e-104;
Matches 318; Conservative 33; Mismatches 21; Indels 1; Gaps 1;

QY 1 MHQAIPQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTTDVEATVNOIKALERVADIYR 60
DB 3 MHQAIPQRKSTRIVYGNVPIGDGAPIAVQSMNTTRTTDVEATVNOIKALERVADIYR 62
QY 61 VSVPTMDAAEAFKLIKQOVNPLVADHFDYRIALKAEGVDCRLNPNIGNEERRIM 120
DB 61 VSVPTMDAAEAFKLIKQOVNPLVADHFDYRIALKAEGVDCRLNPNIGNEERRIM 122
QY 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDQFKVSV 180
DB 121 VVDCARDKNIPIRIGVNASGLEKDOEKYGEPTPOALLBSAMRHVDHLDRLNFDQFKVSV 182
QY 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSGAVSAIGLGLLSEIGDITLRVSL 240
DB 181 KASDVFLAVESYRLAKQIDOPHLGITEAGARSGAVSAIGLGLLSEIGDITLRVSL 242
QY 241 AADPVEIKVGFPIKSLRIRSGINPIACPTCSROEPDVTGTVNALBQRLIEDIITPMV 300
DB 241 AADPVEIKVGFPIKSLRIRSGINPIACPTCSROEPDVTGTVNALBQRLIEDIITPMV 302
QY 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNDMDIQLBARIKASQD 360
DB 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNDMDIQLBARIKASQD 362
QY 360 DEARRIDVOQVEK 372
DB 363 DEKNRIDIKHVEQ 375

RESULT 7
H64063

gcpb protein - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #ext_change 28-Jul-2000
C:Accession: H64063
R:Flotschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

Db 241 AADPVEIKVGFILKSLRIRSRGINFIACPTCSRQGFVIGTVNALRLEDITITPMDV 300
 Qy 301 SIIGCVNNGGEGALVSTLGTGNGKSGLYEDGVRKRLDNNMDIDLEARIKASOLD 360
 Db 301 SIIGCVNNGGEGALVSTLGTGNGKSGLYEDGVRKRLDNNMDIDLEARIKASOLD 360
 Qy 361 EARRIDVOQVEK 372
 Db 361 EARRIDVOQVEK 372

RESULT 2

A91051

hypothetical protein Ecag377 [imported] - Escherichia coli (strain O157:H7, substrain R1
 C:Species: Escherichia coli
 C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
 C:Accession: A91051
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 G:Hayashi, T.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
 A:Reference number: A9629; MUID:2156231; PMID:11258796
 A:Accession: A91051
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-372 <NAV>
 A:Cross-references: GB:BA000007; PIDN:BA936800.1; PID:g13362847; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain R1MD 0509952
 C:Genetics:
 A:Gene: Ecag377
 C:Superfamily: gcpe protein

Query Match 100.0%; Score 1866; DB 2; Length 372;
 Best Local Similarity 100.0%; Pred. No. 1.3e-121;
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHNQAPIQRKSTRIVYGNVPIGDGAPIAVQSMNTRTTDEATVNOIKALERVGADIVR 60
 Db 1 MHNQAPIQRKSTRIVYGNVPIGDGAPIAVQSMNTRTTDEATVNOIKALERVGADIVR 60
 Qy 61 VSVPTMDAAEAFKLIKQVNVPLVADIHFDYRIALKVAEYGVDCRLINPQNGNERIRM 120
 Db 61 VSVPTMDAAEAFKLIKQVNVPLVADIHFDYRIALKVAEYGVDCRLINPQNGNERIRM 120
 Qy 121 VVDCADKNIPIRIGVNASGLEKDOEKYGEPTPOALLSARHVDHLRLNPDQKVS 180
 Db 121 VVDCADKNIPIRIGVNASGLEKDOEKYGEPTPOALLSARHVDHLRLNPDQKVS 180
 Qy 181 KASDVFLAVESYRLAKQIDOPHLGITTEAGARSAGVSAIGLLSEIGDITLRSVL 240
 Db 181 KASDVFLAVESYRLAKQIDOPHLGITTEAGARSAGVSAIGLLSEIGDITLRSVL 240
 Qy 241 AADPVEIKVGFILKSLRIRSRGINFIACPTCSRQGFVIGTVNALRLEDITITPMDV 300
 Db 241 AADPVEIKVGFILKSLRIRSRGINFIACPTCSRQGFVIGTVNALRLEDITITPMDV 300
 Qy 301 SIIGCVNNGGEGALVSTLGTGNGKSGLYEDGVRKRLDNNMDIDLEARIKASOLD 360
 Db 301 SIIGCVNNGGEGALVSTLGTGNGKSGLYEDGVRKRLDNNMDIDLEARIKASOLD 360
 Qy 361 EARRIDVOQVEK 372
 Db 361 EARRIDVOQVEK 372

RESULT 3

E85895

hypothetical protein gcpe [imported] - Escherichia coli (strain O157:H7, substrain EDL93
 C:Species: Escherichia coli
 C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: E85895
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Groetbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apodaca,

Nature 409, 529-533, 2001

A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551

A:Accession: E85895

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-372 <STO>

A:Cross-references: GB:AE005174; NID:g12516909; PIDN:AG57625.1; GSPDB:GN00145; UWGP:237.

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: gcpe

C:Superfamily: gcpe protein

Query Match 100.0%; Score 1866; DB 2; Length 372;
 Best Local Similarity 100.0%; Pred. No. 1.3e-121;
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHNQAPIQRKSTRIVYGNVPIGDGAPIAVQSMNTRTTDEATVNOIKALERVGADIVR 60
 Db 1 MHNQAPIQRKSTRIVYGNVPIGDGAPIAVQSMNTRTTDEATVNOIKALERVGADIVR 60
 Qy 61 VSVPTMDAAEAFKLIKQVNVPLVADIHFDYRIALKVAEYGVDCRLINPQNGNERIRM 120
 Db 61 VSVPTMDAAEAFKLIKQVNVPLVADIHFDYRIALKVAEYGVDCRLINPQNGNERIRM 120
 Qy 121 VVDCADKNIPIRIGVNASGLEKDOEKYGEPTPOALLSARHVDHLRLNPDQKVS 180
 Db 121 VVDCADKNIPIRIGVNASGLEKDOEKYGEPTPOALLSARHVDHLRLNPDQKVS 180
 Qy 181 KASDVFLAVESYRLAKQIDOPHLGITTEAGARSAGVSAIGLLSEIGDITLRSVL 240
 Db 181 KASDVFLAVESYRLAKQIDOPHLGITTEAGARSAGVSAIGLLSEIGDITLRSVL 240
 Qy 241 AADPVEIKVGFILKSLRIRSRGINFIACPTCSRQGFVIGTVNALRLEDITITPMDV 300
 Db 241 AADPVEIKVGFILKSLRIRSRGINFIACPTCSRQGFVIGTVNALRLEDITITPMDV 300
 Qy 301 SIIGCVNNGGEGALVSTLGTGNGKSGLYEDGVRKRLDNNMDIDLEARIKASOLD 360
 Db 301 SIIGCVNNGGEGALVSTLGTGNGKSGLYEDGVRKRLDNNMDIDLEARIKASOLD 360
 Qy 361 EARRIDVOQVEK 372
 Db 361 EARRIDVOQVEK 372

RESULT 4

AB0822

gcpe protein (protein E) [imported] - Salmonella enterica subsp. enterica serovar Typh
 C:Species: Salmonella enterica subsp. enterica serovar Typh
 A:Note: This species has also been called Salmonella typh
 C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

C:Accession: AB0822
 R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
 th, T.; Connerton, P.; Croft, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
 S.; Moutle, S.; O'Gaora, P.

Nature 413, 848-852, 2001
 A:Authors: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
 A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
 A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AB0822

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-372 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD02726.1; PID:g16503738; GSPDB:GN00176

C:Genetics:

A:Gene: STY2768

C:Superfamily: gcpe protein

Query Match 98.1%; Score 1831; DB 2; Length 372;
 Best Local Similarity 98.1%; Pred. No. 3.5e-119;
 Matches 365; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 17:07:18 ; Search time 26 Seconds
(without alignments)
1375.952 Million cell updates/sec

Title: US-09-921-992-50

Perfect score: 1866
Sequence: 1 MHNQAPIQRKSTRIVYGVN.....RAKASQLEARRIDVQVEK 372

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1866	100.0	372	1	S23058	gcpe protein - Esc
2	1866	100.0	372	2	A91051	hypothetical prote
3	1866	100.0	372	2	E85895	hypothetical prote
4	1831	98.1	372	2	AB0822	gcpe protein (prot
5	1622.5	87.0	375	2	AG0350	probable acetyltra
6	1621.5	86.9	376	2	F82283	gcpe protein VC075
7	1603.5	85.9	368	1	H64063	gcpe protein - Hae
8	1372.5	73.6	371	2	F83171	conserved hypotet
9	1282	68.7	368	2	E84563	gcpe protein (impo
10	1042.5	55.9	383	2	H87354	gcpe protein (impo
11	887	47.5	349	2	G97121	gcpe protein (impo
12	882.5	47.3	367	2	A83825	peptidoglycan acet
13	868	46.5	368	2	A11354	probable peptidogl
14	859.5	46.1	377	2	C69955	hypothetical prote
15	828	44.4	357	2	F70433	gcpe protein - Aqu
16	815.5	43.7	384	2	T35407	conserved hypotet
17	794	42.6	359	1	A64598	protein E - Helico
18	790	42.3	359	2	E71914	hypothetical prote
19	751.5	40.3	387	2	F70886	probable gcpe prot
20	712.5	38.2	392	2	G87106	conserved hypotet
21	706	37.8	344	2	D72321	gcpe protein - The
22	690	37.0	357	2	A81339	gcpe protein homol
23	618.5	33.1	403	2	S77159	gcpe protein - Syn
24	613.5	32.9	408	2	AF2118	hypothetical prote
25	588	31.5	404	2	F71324	probable gcpe prot
26	543.5	29.1	429	2	D75526	gcpe protein - Del
27	527	28.2	421	2	D81098	conserved hypotet
28	527	28.2	421	2	H81643	gcpe protein (impo
29	509	27.3	437	2	AH3285	gcpe protein (impo

30	504.5	27.0	601	2	C81715	gcpe protein TC032
31	492	26.4	416	2	AB2911	peptidoglycan acet
32	492	26.4	440	2	H97685	hypothetical prote
33	487	26.1	602	2	E71562	hypothetical prote
34	470.5	25.2	613	2	C86537	gcpe protein (impo
35	470.5	25.2	613	2	E72087	gcpe protein CP038
36	459	24.6	417	2	B82542	conserved hypotet
37	126	6.8	533	2	S76101	hypothetical prote
38	114	6.1	524	2	H69099	conserved hypotet
39	112.5	6.0	686	2	H90226	minichromosome mai
40	111	5.9	486	2	A75045	2-isopropylmalate
41	108	5.8	690	2	E72337	translation initia
42	106.5	5.7	284	2	D86701	mathionine aminope
43	106.5	5.7	531	2	S52294	2-isopropylmalate
44	106.5	5.7	531	2	AH2410	2-isopropylmalate
45	106	5.7	748	2	A64319	carbon-monoxide de

ALIGNMENTS

```
RESULT 1
S23058
gcpe protein - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #ext_change 01-Mar-2002
C/Accession: S23058; B65028
R/Baker, J.; Franklin, D.B.; Parker, J.
FEMS Microbiol. Lett. 94, 175-180, 1992
A/Title: Sequence and characterization of the gcpe gene of Escherichia coli.
A/Reference number: S23057
A/Accession: S23058
A/Molecule type: DNA
A/Residues: 1-372 <BAK>
A/Cross-references: EMBL:X64451, NID:G41540, PID:G41542
A/Experimental source: strain K-12
R/Battner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Coj
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A/Title: The complete genome sequence of Escherichia coli K-12.
A/Reference number: A64720; MUID:97426517; PMID:9278503
A/Accession: B65028
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-372 <BAK>
A/Cross-references: GB:A6000338, GB:U00096, NID:G1788862, PIDN:AAC75568.1; PID:G1788863;
A/Experimental source: strain K-12, substrain MG1655
C/Genetics:
A/Gene: gcpe
C/Superfamily: gcpe protein

Query Match 100.0%; Score 1866; DB 1; Length 372;
Best Local Similarity 100.0%; Pred. No. 1.3e-12;
Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHNQAPIQRKSTRIVYGVNPIGDAPIAVQSMTRTDTVATNOIKALRVGADYR 60
DB 1 MHNQAPIQRKSTRIVYGVNPIGDAPIAVQSMTRTDTVATNOIKALRVGADYR 60
QY 61 VSEPTMDAAEAKFLIKQGVNVPVLVADIHFDYRIALKVAEYGVDCRLINPQNGEERIRM 120
DB 61 VSEPTMDAAEAKFLIKQGVNVPVLVADIHFDYRIALKVAEYGVDCRLINPQNGEERIRM 120
QY 121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPOLLESAMRHVDHLDRLNFQPKVSU 180
DB 121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPOLLESAMRHVDHLDRLNFQPKVSU 180
QY 122 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPOLLESAMRHVDHLDRLNFQPKVSU 180
DB 122 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPOLLESAMRHVDHLDRLNFQPKVSU 180
QY 181 KASDVFVAESYRLAKQIDQPLHIGTEAGAGAGAKSAIGGLLSEGIQDPLRVSL 240
DB 181 KASDVFVAESYRLAKQIDQPLHIGTEAGAGAGAKSAIGGLLSEGIQDPLRVSL 240
QY 241 AADPVEIKVGFIDILKSRIRSGINFIACPTCSQOEFDVIGTVNALQORLEDITPMQV 300
DB 241 AADPVEIKVGFIDILKSRIRSGINFIACPTCSQOEFDVIGTVNALQORLEDITPMQV 300
```

CC are used: (i) to increase the isoprenoid levels in viruses and cells;
CC (ii) for determining the enzymatic activity of gcpe and yf8b proteins;
CC and (iii) to identify compounds that inhibit activity of gcpe, i.e.
CC potential antibacterial, antimycotic, antiparasitic or antiviral agents
CC for use in humans or animals, or antiviral, antiparasitic, fungicidal
CC or herbicidal agents for agriculture.

XX Sequence 679 AA;

Query Match 24.7%; Score 460; DB 22; Length 679;

Best Local Similarity 38.9%; Pred. No. 2.3e-37;
Matches 112; Conservative 57; Mismatches 81; Indels 38; Gaps 8;

QY 9 RKKSTR-IYGNVPIGDGAPIAVQSMNTTRTVDVATVNOIKALEVAGDIYVSVPTMD 67
DB 117 KRPTREVVIGNVKIGNNKRIAIQTWASCDTRNVECVQIRKCKDGLADIVRLTVQGVQ 176
QY 68 AAEAF-----KLIKOQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN----- 114
DB 177 EAQASYHIKELKLSNNVNIPLVADIHFNPKIALMADV-FEKIRVNGVYVDRKKWIDK 235
QY 115 -----EERIRMVDCARDKNIPIRIGVAGSLKDLQEKYGEPTPOALL 158
DB 236 VYKTEEFDEGKLFIEKEFVPLIEKCKR-LNRAIRIGTHGSLSSVLSYGD-TPLATV 293
QY 159 ESAMRHVDHLDRNFPQFVSVASDVFLAVESYRL-AKQIDQ----PLHIGITTAGCA 213
DB 294 ESAMRFSDLCENNFNPLVFSMKASNAVYMIQSYRLVSKOYERNMMPFIHLGVTEAGFG 353
QY 214 RSGAVSAIGLGLLSEIGDTRLVSLADPVEIKVGFILKSLRIR 261
DB 354 DNGRIKSYLGISLVDIGDITRISLTEDPWEELTPCKKLVENLKKR 401

RESULT 15

AAW90883
ID AAW90883 standard; Protein; 752 AA.

XX AAW90883;

XX 07-JUL-2000 (first entry)

DE P. falciparum gcpe protein.

XX 1-deoxy-D-xylulose-5-phosphate reductoisomerase; isoprenoid biosynthesis;
KW deoxy-D-xylulose-phosphate; 1-deoxy-D-xylulose-5-phosphate synthase;
KM gcpe; antimycotic; antibiotic; antiviral; antiparasitic; antimicrobial;
XX bactericide; fungicide; herbicide; treatment; disease.

XX Plasmodium falciparum.

XX WO200017233-A2.

XX 30-MAR-2000.

XX 22-SEP-1999; 99WO-EP07055.

XX 22-SEP-1998; 98DE-1043279.

XX 21-MAY-1999; 99DE-1023567.

XX (JOMAA/) JOMAA H.

XX Jomaa H;

XX WPI; 2000-283543/24.

XX N-PDB; AAX82996.

PT New nucleic acid, useful therapeutically and to screen for e.g.
PT antimicrobials and herbicides, encode proteins involved in isoprenoid
PT biosynthesis by the deoxy-D-xylulose-phosphate route

PS Claim 3; Page 41-44; 44pp; German.

CC This invention describes novel polypeptides and their analogs which are
CC involved in isoprenoid biosynthesis by the deoxy-D-xylulose-phosphate
CC route. The proteins described include 1-deoxy-D-xylulose-5-phosphate
CC reductoisomerase, 1-deoxy-D-xylulose-5-phosphate synthase and the gcpe
CC protein. The products of the invention have antimycotic, antibiotic,
CC antiviral, antiparasitic, antimicrobial, bactericide, fungicide and
CC herbicide activity. The encoding nucleic acid sequences of the invention
CC are used for prevention or treatment of diseases in humans and animals
CC and also to screen compounds enzyme inhibitory activity (potential
CC antimicrobials, antiparasitic agents, antivirals, fungicides,
CC bactericides and herbicides, for use in human or veterinary medicine or
CC agriculture). This sequence represents the Plasmodium falciparum gcpe
CC protein described in the method of the invention.

XX Sequence 752 AA;

Query Match 23.9%; Score 446; DB 21; Length 752;

Best Local Similarity 38.2%; Pred. No. 7.1e-36;
Matches 110; Conservative 56; Mismatches 84; Indels 38; Gaps 8;

QY 9 RKKSTR-IYGNVPIGDGAPIAVQSMNTTRTVDVATVNOIKALEVAGDIYVSVPTMD 67
DB 117 KRPTREVVIGNVKIGNNKRIAIQTWASCDTRNVECVQIRKCKDGLADIVRLTVQGVQ 176
QY 68 AAEAF-----KLIKOQVNVPLVADIHFDYRIALKVAEYGVDCLRINPGNIGN----- 114
DB 177 EAQASYHIKELKLSNNVNIPLVADIHFNPKIALMADV-FEKIRVNGVYVDRKKWIDK 235
QY 115 -----EERIRMVDCARDKNIPIRIGVAGSLKDLQEKYGEPTPOALL 158
DB 236 VYKTEEFDEGKLFIEKEFVPLIEKCKR-LNRAIRIGTHGSLSSVLSYGD-TPLGMY 293
QY 159 ESAMRHVDHLDRNFPQFVSVASDVFLAVESYRL-AKQIDQ----PLHIGITTAGCA 213
DB 294 ESAMRFSDLCENNFNPLVFSMKASNAVYMIQSYRLVSKOYERNMMPFIHLGVTEAGFG 353
QY 214 RSGAVSAIGLGLLSEIGDTRLVSLADPVEIKVGFILKSLRIR 261
DB 354 DNGRIKSYLGISLVDIGDITRISLTEDPWEELTPCKKLVENLKKR 401

Search completed: November 23, 2003, 17:15:04
Job time : 53 secs

Db 258 GIIKSAVIGITLAEGLDITRCSLTGCTFEIRVDCSLRHRTX 302

RESULT 13

AAE19650 standard; Protein; 740 AA.

AAE19650;

31-MAY-2002 (first entry)

Arabidopsis thaliana GCPE protein.

GCPE gene; methyl-D-erythritol phosphate; MEP; dimethylallyl diphosphate; transgenic plant; isoprenoid compound; tocopherol; isopentyl diphosphate; food; feed source; transformation; single nucleotide polymorphism; SNP; oxidative stress tolerance; UV tolerance; transformation; chromosome V; plant; GCPE protein.

Arabidopsis thaliana.

Key Location/Qualifiers

FT Cleavage-site 32..33 /note="Plastidial targeting peptide cleavage site"

MO200212478-A2.

14-FEB-2002.

06-AUG-2001; 2001WO-US24335.

07-AUG-2000; 2000US-223483P.

(MONS) MONSANTO TECHNOLOGY LLC.

Boronat A, Campos N, Rodriguez-concepcion M, Rohner M, Seeman M;

Valentin HE, Venkatesh TV, Venkatramesh M;

WPI: 2002-227151/28.

N-PSDB; AAD31201.

GCPE nucleic acid which is an essential gene of the methyl-D-erythritol phosphate pathway, encoding a fully defined GCPE protein which is

useful for increasing levels of tocopherol substrates in plants -

Claim 1; Page 140-142; 155pp; English.

The invention relates to gcpe nucleic acid molecule, an essential gene of methyl-D-erythritol phosphate (MEP) pathway) that encodes rice, Arabidopsis thaliana or Escherichia coli GCPE protein. gcpe is useful for producing a transgenic plant such as Brassica campestris, B. napus, canola, castor bean, coconut, cotton, crambe, linseed, maize, mustard, oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower, or wheat with an increased isoprenoid (tocopherol) compound level. The expression of GCPE protein in organisms increases the level of tocopherol substrate such as isopentyl diphosphate and dimethylallyl diphosphate biosynthesis. Transgenic organisms overexpressing GCPE protein can nutritionally enhance food and feed sources. Overexpression of GCPE protein in transgenic plant may provide tolerance to stresses e.g., oxidative stress tolerance such as to oxygen or ozone, UV tolerance, etc. GCPE may be used to obtain nucleic acid molecules from the same species, and to obtain nucleic acid molecules from used as or primers. The recombinant vectors are used in plant transformation or transfection. gcpe can also act as markers capable of detecting polymorphisms such as single nucleotide polymorphisms (SNPs). GCPE is also used to determine the level or pattern of expression of the protein. The present sequence is Arabidopsis thaliana GCPE protein. gcpe gene is located on chromosome V.

Sequence 740 AA;

Query Match 24.8%; Score 462; DB 23; Length 740;

Best Local Similarity 38.6%; Pred. No. 1,6e-37; Matches 105; Conservative 47; Mismatches 86; Indels 34; Gaps 5;

QY 9 RKSTRIYVGNVPIGDPAPVAGSMNTRTTVDVATVNOIKALERVAGADIVRSVPTMDA 68

Db 84 RKRTRVWGNVAGLSEHPRIQITMTSDTKITGVDEVWRIADKAGADIVRTVQGRKE 143

QY 69 AEAFA-----KLIRQVNVPLVADIFDFYRIALVYAEVGVDCLRINFGNIGNERIRRVVD 123

Db 144 ADACFEIKDVLVOLANNIPVLADIHFAPTVALVAEC-PKIRVNGNFPADRAQPFETID 202

QY 124 CARDK-----NIPIRIGVAGSLDKDLQEKYGEPTQALLLSA 161

Db 203 YTEDEVQKELQHTIEQVFTPLVECKKYGKRAMRIGTNHGSLSDRIMSYGD-SPRGWESA 261

QY 162 MRVVDHLDRINFPQFVKVSVKASVFLAVESYRIILAOI-----DQPLHIGITAGARSG 216

Db 262 FEPARICRKLIDYNNFVFSMKASNPVIMVOAYRLLVAMVGVHGWDPYPLHGVTEAGEGEGD 321

QY 217 AVKSAIGLGLLSEGIQDTLRVSLADPVEEI 248

Db 322 RKSATIGITLDDGIDTIRVSLTEPPEEI 353

RESULT 14

AAE45693 standard; Protein; 679 AA.

AAE45693;

15-MAR-2001 (first entry)

P. falciparum gcpe protein.

Isoprenoid; gcpe; yfgb; antimicrobial; transgenic plant; agriculture; antimycotic; antiparasitic; antiviral; fungicidal; herbicidal.

Plasmodium falciparum.

MO200072022-A1.

30-NOV-2000.

20-MAY-2000; 2000MO-EP04592.

PR 21-MAY-1999; 99DE-1023567.

PR 21-MAY-1999; 99DE-1023568.

(JOMA/) JOMAA H.

Jomaa H;

WPI: 2001-025196/03.

N-PSDB; AAC82654.

Incorporating gcpe and yfgb genes into viruses and cells, for increasing isoprenoid content and identifying e.g. antimicrobial agents, comprises using DNA sequences from bacteria or parasites -

Disclosure; Page 21-23; 36pp; German.

This invention describes a novel method for incorporating gcpe and yfgb genes into viruses and cells for increasing isoprenoid content and identifying e.g. antimicrobial agents, comprising using DNA sequences (I) from the gcpe or yfgb genes of bacteria or parasites or DNA sequences (II) which hybridize to the specified genes or encode a plastid protein with the same biological activity as those encoded by the genes. The invention also describes (1) plant cells containing (I) or (II); (2) transformed plant cells, and transgenic plants regenerated from them, that contain (I) or (II); (3) determining the enzymatic activity of a gcpe protein; or (4) screening compounds (A) that have antimycotic, antiparasitic or antiviral activity in humans or animals or antiviral, antiparasitic, fungicidal or herbicidal activity in plants. (I) and (II)

XX 27-NOV-1998; 98WO-IB01939.
 XX
 PR 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97FR-0015041.
 PR 17-DEC-1997; 97FR-0016034.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R;
 XX
 DR WPI, 1999-371125/31.
 XX
 PS Genome sequence of Chlamydia trachomatis
 Disclosure; Page 943-944; 1755pp; English.
 XX
 CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nongonococcal urethritis,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perihepatitis, bartolinitis; pneumonia in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.
 XX
 SQ Sequence 611 AA;

Query Match 26.3%; Score 490; DB 20; Length 611;
 Best Local Similarity 23.9%; Pred. No. 1,7e-40;
 Matches 134; Conservative 76; Mismatches 109; Indels 242; Gaps 8;

QY 9 RRKSTRIYGVNPIGDCAPIAVOSMTNRTTVEATVNCIKALERVGADIVRVSVPTMDA 68
 DB 20 RRKTLPRIGIDLFGVSEHSIKISMTTATTVDGTVRQICALQELGCDIVRVTVQGLRE 79
 QY 69 AEF-----KLIIQVNVPLVADIHFDYRIALVAEYGVDCIINPN-----1 112
 DB 80 VHAECHEIKORLIQONISIPVADIHFFPPQAHVVDV-VKVRINPNVYDKRMFTGKI 138
 QY 113 GNEERIRMVVDCARDKNIP-----RIGNVAGSLEKLOEKYGEPTPOLLES 161
 DB 139 YSBEQVAHSLBHLMKRSPLEVECKRLGKMRIGVNHGSLSEKVTQRYN-TIEGMVYSA 197
 QY 162 MRHVHDLRLNFDQFKVSVKASDVFLAVESYRLAKOIDO-----PLHLGITTEAGARS 216
 DB 198 LEVAVCVAMDYHDVIFSMKSNPKVVAAYRSIAVELDQREMSYPLHLGVTTEAGSGTAG 257
 QY 217 AVKSAIGLILLSEIGDITLRVSLADPVEIKV-----250
 DB 258 IVKSAVGIITLSEGLDITRCSLPGINEIPICIDLLKOTTELSEKMGADNPAIHS 317
 QY 251 -----250
 DB 318 SKOLGTRNTLTPPMNVVGLINLNDVOLLTAPEIELOCLGIDITTKIDPTTEGAV 377
 QY 251 -----250
 DB 378 VPKAMSSPIVSEIEKHLVFNKEDAPILNPMNEEWLSEETLSAPFVPEVTDIHARR 437
 QY 251 -----250
 DB 438 FSLRHSSTOPVCLSFSLDHLKNEAIIIDLSARLALLDGLSCVLLDPVDIKLSRTL 497
 QY 251 GFDILKSLIRSRGINFIACPTCSROEFDVIGTVALLEORLEDIITPMDSIIGCVVNGP 310
 DB 498 GFDILGASNRSTVTEVSCPCGCRITLFDLLAVSQRIRERTKHLPGSLAKIAVWGCIYNGP 557
 QY 311 GEALVSTLVGTGANKSGLYE 331

DB 558 GENADADFGYVG--SKPGMID 576
 RESULT 12
 ID AAY34971
 ID AAY34971 standard; Protein; 621 AA.
 XX
 AC AAY34971;
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE Amino acid sequence of a Chlamydia pneumoniae protein.
 XX
 KM Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KM sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KM vaccine; neutralising epitope.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN MO9927105-A2.
 XX
 PD 03-UTN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R;
 XX
 DR WPI, 1999-357842/30.
 XX
 PS Genome sequence of Chlamydia pneumoniae
 PT
 Page 885-886; Disclosure; 1912pp; English.

XX AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAY91990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 XX
 SQ Sequence 621 AA;

Query Match 25.2%; Score 470.5; DB 20; Length 621;
 Best Local Similarity 38.9%; Pred. No. 1,7e-38;
 Matches 111; Conservative 55; Mismatches 84; Indels 35; Gaps 5;

QY 9 RRKSTRIYGVNPIGDCAPIAVOSMTNRTTVEATVNCIKALERVGADIVRVSVPTMDA 68
 DB 20 RRKTLPRIGIDLFGVSEHSIKISMTTATTVDGTVRQICALQELGCDIVRVTVQGLRE 79
 QY 69 AEFKLIKQO-----VNVPLVADIHFDYRIALVAEYGVDCIINPN-----111
 DB 80 AQACEKIKERLIALGINIPVADIHFFPPQAMLVADF-ADKVRINPNVYDKRMFTGKR 138
 QY 112 -----IGNEERIRMVVDCARDKNIPIRIGNVAGSLEKLOEKYGEPTPOLLES 160
 DB 139 IYTESVYASQSLRLEKEFPALVECKRLGKMRIGVNHGSLSERIMQYGD-TIEGMVAS 197
 QY 161 AMRHVHDLRLNFDQFKVSVKASDVFLAVESYRLAKOIDO-----QPLHLGITTEAGARS 215
 DB 198 ALBYAVCEKLNRDVVFMSKSNPKIMVATARQLKDDJANGWLPLHLGVTTEAGMGVD 257
 QY 216 GAVKSAIGLILLSEIGDITLRVSLADPVEIKVGFILKSLRI 260

CC encoded by ABZ71062 to ABZ71130 (II), which are isolated from
CC *Mycobacterium tuberculosis*. (I) are encoded by genes (II) whose
CC expression is induced or up-regulated during culture of a mycobacterium
CC under conditions defined by a dissolved oxygen tension of at least 10%
CC air saturation measured at 37 plus degrees Celsius, when compared with
CC dissolved oxygen tension of at least 40% air saturation measured at 37
CC plus degrees Celsius. (I) and (II) have antibacterial, tuberculostatic
CC and immunostimulant activities, and can be used in vaccines and gene
CC therapy. (I) and (II) can be used for the manufacture of a medicament
CC for treating or preventing a mycobacterial infection. They can also be
CC used for the manufacture of a diagnostic reagent for identifying a
CC mycobacterial infection.

Sequence	387	AA;
----------	-----	-----

Query Match	40.3%	Score 751.5	DB 24	Length 387
Best Local Similarity	42.7%	Pred. No. 2e-67		
Matches 156	Conservative 64	Mismatches 134	Indels 11	Gaps 2

QY	9	RRKSTETIYVNGPIDGAPILAVOSMTNRTTGVETALNQIALERVAGDIDYRVSVPIMDA	68
		18 RRAITROLAMGNVGVSDHPSVOSMCTTKTHDVNSTLOOLIELTRAGCDIDYRACPROED	77
QY	69	AEAFLKIKOOVWVPLVADHPDYRIALKVAEYGVCLIRINEGNIGN--EERIRMTVCARD	127
Db	78	ADLARIARHSQIPVAVDHPQRIYFAIDMGCAAVVNGNIKEPGRGVEFAKXAGA	137
QY	128	KNIPIRIGVAGSLEKDLQEKYGEPTPOALLSESARHVDHLDRINFDQFKVSVKASDVFL	187
Db	138	AGIPIRIGVAGSLDRKFMEEKYKATPEALVESALMEASLFEENGFGDIKITSVGNPDV	197
QY	188	AVESYRLKAKJDOPLHGITETAGARGASVNSAIGLULLSEGIQPTLRVSLAADVEE	247
Db	198	MVAAYELLARCDYPLHGLVTEGPRPQOTIKSAVAFALLSRGIGDTIRVLSAPVEE	257
QY	248	IKVGFIDILKSLIRSRGINFIACPTCSRQEPFVIGTVALERQRLIEDITTPMDVSIICGV	307
Db	258	VKKGNGVLESIMLRPSLEIVSCPSCGRAQVDVYTLAMEYVARGLDGIDVPLRAVAMCGV	317
QY	308	NGRGEALVSTLGVTGNGKKSGLYEDGVKRDLINDNDMIDOLEARIKAKASQDEARIDV	367
Db	318	NGGGEAREADLGVASNGKGOIFVIRGEVILKTIPEAQIVETL-----IEEAMRLTAA	367
QY	368	QQVEK	372
Db	368	EMGEQ	372

ID	ABP78786 standard; Protein; 421 AA.
XX	
XX	ABP78786;
DT	07-MAR-2003 (first entry)
XX	
DE	N. gonorrhoeae amino acid sequence SEQ ID 4102.
XX	
KW	Antibacterial; infection; vaccine; gene therapy
XX	
XX	
OS	Neisseria gonorrhoeae.
XX	
PN	WO200279243-A2.
XX	
PD	10-OCT-2002.
XX	
PF	12-FEB-2002; 2002WO-IB02069.
XX	
PR	12-FEB-2001; 2001GB-0003424.
XX	
XX	
PA	(CHIR-) CHIRON SPA.
XX	
TI	Fontana MR, Pizza M, Masignani V, Monaci E;

XX WPI; 2003-058415/05.
DR
DR N-PSDB; ABZ39756.
YY

PT New protein from *Neisseria gonorrhoeae*, useful for the manufacture of a
medicament for treating or preventing *N. gonorrhoeae* infection -
XX
PS Disclosure; Page 487; 815pp; English.

CC The present invention relates to proteins from *Neisseria gonorrhoeae*.
CC Also disclosed are the nucleic acid molecules encoding the proteins and
CC antibodies that specifically bind to the proteins. The composition
CC comprising the protein, nucleic acid or antibody is useful for the
CC manufacture of a medicament for treating or preventing *N. gonorrhoeae*
CC infection, this may be in the form of a vaccine or gene therapy.
CC Sequences given in records ABP76736-ABP81046 represent nucleic acid
CC molecules of the invention.

SQ Sequence 421 AA;

Query Match	28.6%	Score	534	DB	24	Length	421
Best Local Similarity	32.3%	Pred. No.	3.1e-45				
Matches	135	Conservative	87	Mismatches	14	Indels	54
						Gaps	12

QY		7	IORRKTIRVYGAVPLRGDAPLVAOSMNTRTTDEAVNNOIKALERVAGAIIVAVSPTM	66
Dd		4	LORRKHÖVLITHITVGSBAPVVIOSMTNTTDADKATALOIKEISDAGSEWRIITVNSP	63
QY		67	DAAEAFKLIOOVN-----VPLVADIHP-VRIALKVAEYG--VDICRINGING-----N	114
Dd		64	EAAASKVAEIRRRLDDMGVATPLIGPFHNGERILLAPPEGKALKSRINGNVGKVG	122
QY		115	EEERIRVVUCCADKKNIPIRIGNAGSLEKIDOEK-----GEPT-----QALLESA	167
Dd		124	DEFAPAMITTAENDKAVRIGVMWSLIQSIAKRMMDANLVSAKPREEVMEKALITVA	183
QY		162	MRRVHDLRNLNPOEKV---SVKASVPFLAVESYRLAKOIDQPALGITTEAGSGARSAVK	219
Dd		184	LSBAEAVALILGPEDKIILSCVSAVHDLIOVRELSGRVCYPPLHLGITTEAGMSKGITVA	243
QY		220	SAIGGLILLSBIGIDTLRVSLAADP-----VEEIKVGPILKSIRRSKGINFIACPTCSR	275
Dd		244	STAAISVLLQEBIGDTIRISLTLPBEGSPRGTEVVVGGEIIQTGMGRSPETPMVTACPCGCR	309
QY		276	QEEDVYIGTV-----NALBOR-----LEDITTPMDVSIIIGCVNVGPGBALVSTLGV----	320
Dd		304	TTSJVFQEOELADVOUYLTRÖKMISMWTLTYLPAGESINVAVMGVNNGPESGIADIGISLPG	363
QY		321	TGGNKSKGLYEYGVKDRBLDNNDMDOLUARIR-----AKSQULDEARRIDIYOQV	370
Dd		364	TGETPVAPAYVDGERKVTLTKGNINLAISEPLAIVEEVYKTNVGNSSKKNKGVAPIQSL	421

	RESULT 11
AAV37164	
ID	AAV37164 standard; Protein; 611 AA.
XX	
AC	AAV37164;
XX	
DT	07-OCT-1999 (first entry)
XX	
DE	Amino acid sequence of a Chlamydia trachomatis protein.
XX	
KM	Vaccine; eye disease; conventional trachoma; nonendemic trach
KM	paratrachoma; inclusion conjunctivitis; genital disease; peritri
KM	nongonococcal urethritis; epididymitis; cervicitis; salpingitis
KM	bartolinittis; pneumopathy; venereal lymphogranulomatosis.
XX	
OS	Chlamydia trachomatis.
XX	
PN	MO9928475-A2.
XX	
DD	10-JUN-1999.

QY 69 AAEFKLIKQGVNPLVADHFDYRIALKAAYGVDCIRINPIGN-EEIRIMVUDCARD 127
 DB 65 AEAFLPIIAKSPPIADIHFQPKYIFPAIDGCAAVRVPNGIKERPDGVKAVAAAGD 124
 QY 128 KNPIPIGVNAGSLEKDLOEKY-GEPIPPALLSNAHVDHLRNLNFDQKVSXKASDVF 186
 DB 125 AGPIPIGVNAGSLEKDLOEKY-GEPIPPALLSNAHVDHLRNLNFDQKVSXKASDVF 184
 QY 187 LAVESYRLAKQIDOPHLGITEAGARSGAVKSAIGLLSEGIGDTLRVSLADPVE 246
 DB 185 LMEVATRLQAEQSDYPLHLGVTENGPKEFTIKSVAFGLLSQIGDTLRVSLADPVE 244
 QY 247 EIKVGFILKSLIRSGINFIACPTCSROEPVYIGTVNALBEQLEDTITPMDSIIGCV 306
 DB 245 EIKVGFILKSLIRSGINFIACPTCSROEPVYIGTVNALBEQLEDTITPMDSIIGCV 304
 QY 307 VNGPGEALVSTLGVTKGNKSGLYEDGVKRDLDNNMDIQLEARIKAKASQLEARRI 365
 DB 305 VNGPGEALVSTLGVTKGNKSGLYEDGVKRDLDNNMDIQLEARIKAKASQLEARRI 363

RESULT 8

AA81205
 ID AAG81205 standard; Protein; 387 AA.

AA81205;
 XX

04-SEP-2001 (first entry)
 XX

Mycobacterium tuberculosis potential drug target protein SEQ ID 256.
 XX

Drug target; growth; organism viability; characterisation.
 XX

Mycobacterium tuberculosis.
 XX

MO200135317-A1.
 XX

17-MAY-2001.
 XX

13-NOV-2000; 2000MO-US31152.
 XX

12-NOV-1999; 99US-0165086.
 XX

12-NOV-1999; 99US-0165124.
 XX

01-FEB-2000; 2000US-0179531.
 XX

(REGC) UNIV CALIFORNIA.
 XX

Eisenberg D, Rotstein SH, Marcotte EM,
 XX

WPI; 2001-329193/34.
 XX

N-PSDB; AAHS2056.
 XX

Identifying nucleotide or polypeptide sequence for use as drug target,
 PT involves providing algorithm that analyzes a functional relationship
 PT between nucleotide or polypeptide sequences, and comparing the
 PT sequences -

Disclousure; Page 182; 207pp; English.
 PS

XX This invention relates to a method for identifying a nucleotide or
 CC polypeptide sequence that may be a drug target, or essential for growth
 CC or viability of an organism. Polynucleotide sequences AAHS1947 - AAHS2092
 CC represent DNA encoding proteins AAG81096 - AAG81241. Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analyzing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism.
 CC

SQ Sequence 387 AA;

Query Match 40.3%; Score 751.5; DB 22; Length 387;

Best Local Similarity 42.7%; Pred. No. 26-67;

Matches 156; Conservative 64; Mismatches 134; Indels 11; Gaps 2;

QY 9 RKRSTRIYVNVPIGDGAPIAVQSMNTTRTVEATVNOIKALERVGADIVRSVPTMDA 68
 DB 18 RATRQLVMVGNVGVSDHVSQVSMCTTKTHDVNSTLQIAELTAAACDIVRVACPRQED 77
 QY 69 AAEFKLIKQGVNPLVADHFDYRIALKAAYGVDCIRINPIGN-EEIRIMVUDCARD 127
 DB 78 AALAEIASHSQIPVADHFDYRIALKAAYGVDCIRINPIGN-EEIRIMVUDCARD 137
 QY 128 KNPIPIGVNAGSLEKDLOEKYGEPTPPALLSNAHVDHLRNLNFDQKVSXKASDVF 187
 DB 138 AGPIPIGVNAGSLEKDLOEKYGEPTPPALLSNAHVDHLRNLNFDQKVSXKASDVF 197
 QY 188 AAVSYRLAKQIDOPHLGITEAGARSGAVKSAIGLLSEGIGDTLRVSLADPVE 247
 DB 198 MVAAYVELLARCQDYPHLGVTENGPKEFTIKSVAFGLLSQIGDTLRVSLADPVE 257
 QY 248 EIKVGFILKSLIRSGINFIACPTCSROEPVYIGTVNALBEQLEDTITPMDSIIGCV 307
 DB 258 VIKVGVNVLSTLGVTKGNKSGLYEDGVKRDLDNNMDIQLEARIKAKASQLEARRI 317
 QY 308 NGPGEALVSTLGVTKGNKSGLYEDGVKRDLDNNMDIQLEARIKAKASQLEARRI 367
 DB 318 NGPGEALVSTLGVTKGNKSGLYEDGVKRDLDNNMDIQLEARIKAKASQLEARRI 367
 QY 368 QOVER 372
 DB 368 EMGEQ 372

RESULT 9

ABP57500
 ID ABP57500 standard; Protein; 387 AA.

ABP57500;
 XX

28-APR-2003 (first entry)
 XX

Mycobacterium tuberculosis protein SEQ ID NO:129.
 XX

Mycobacterium tuberculosis; infection; antibacterial; tuberculosis;
 KW immunostimulant; vaccine; gene therapy; mycobacterial infection.
 XX

Mycobacterium tuberculosis.
 OS

WO2003000721-A2.
 XX

03-JAN-2003.
 XX

21-JUN-2002; 2002MO-GB02845.
 XX

22-JUN-2001; 2001GB-0015365.
 XX

07-SEP-2001; 2001GB-0021780.
 XX

(MICR-) MICROBIOLOGICAL RES AUTHORITY.
 PA

James BW, Bacon J, Marsh P;
 XX

WPI; 2003-201403/19.
 XX

N-PSDB; ABZ71126.
 XX

New mycobacterial peptide, its fragment, variant or derivative, useful
 PT as vaccine for treating or preventing mycobacterial infections, and as
 PT diagnostic reagents for identifying such infections -

Claim 2; Page 235-236; 246pp; English.
 PS

ABP57436 to ABP57504 represent mycobacterial amino acid sequences (1)
 CC

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
XX dermatological; osteopathic; neuroprotectant.
OS Propionibacterium acne.
XX
XX MO200181581-A2.
XX
XX 01-NOV-2001.
XX
XX 20-APR-2001; 2001MO-US12865.
XX
XX 21-APR-2000; 2000US-199047P.
XX 02-JUN-2000; 2000US-208841P.
XX 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'atsomeuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI: 2001-616774/71.
XX N-PSDB; AASS9573.
XX
XX Propionibacterium acne polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX Example 1; SEQ ID No 17407; 10699P; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acne immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acne. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acne is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acne in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acne proteins. These antibodies can be used to
CC downregulate expression and activity of P. acne polypeptides and
CC therefore treat P. acne infections. The antibodies may also be used as
CC diagnostic agents for determining P. acne presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 415 AA:
SQ
Query Match 42.1%; Score 785.5; DB 22; Length 415;
Best Local Similarity 44.2%; Pred. No. 7.9e-71;
Matches 159; Conservative 68; Mismatches 132; Indels 1; Gaps 1;
QY 9.RRKSTIYGVNPIGCGAPPAVOSMTNRTTDEATVNOIKALERVGADIVRSVPTMDA 68
DB 53.RRTHOIKVGDVAVGGAPISVOSMTTKTHNDGATLQOIALTAAGCDIVRAVACPTDXD 112
QY 69.AEAFKLIKQOVNPLVADHIFDVRIALKVAEYGVDCRINPGIGN-EERIIRNVVCCARD 127
DB 113.AEVLPIIAKSGQIPVADHIFQPKYVFQALIEAGCGARVAVPGIRKFPDDISICQAAE 172
QY 128.KNPIRIGVAGSLIEDLOEKYPTPOALLSASMRHVHDLRLNFDQFVSVKASDVPL 187
DB 173.HGHSIITGVNAGSLDRILDKGAPPAEAMVESALMEASLFEQYGRDFRISVXKHDPV 232
QY 188.AVESYLLAKQIDOPHLGITEGARGSAVKSAGLGLLSGIGDITLRSVLAADPVEE 247
DB 233.MIRAVEQLAACDYPHLGVTGEGAPFOGTRKSAVAFGHLLAEGIDITRVSISADPVEE 292

QY 248.IKVGFILKSLRISRGINFACPTCSRQEFVIGTVALLEORLEDIITPMDSIIGCVV 307
DB 293.VKVGIKILSLNLRPRGLIEVSCPCRCQVDVLTLANDVTALEGIDAPLRAVAVGCVV 352
QY 308.NGPGNALVSTLGTGTGANKKSGLYEDGVKRDRLDNNMDIQLEARIKAQSOLDPEARIDV 367
DB 353.NGIEGREGADLVGAAGNKGKIFKHEGVIIRTVPEGEIVQFLVGEANRMADEMDDTGAVEV 412
RESULT 7
AAG91951
ID AAG91951 standard; Protein, 378 AA.
XX
XX AAG91951;
AC
XX 26-SEP-2001 (first entry)
DT
XX C glutamicum protein fragment SEQ ID NO: 5705.
DE
XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis.
XX
XX Corynebacterium glutamicum.
OS
XX
XX BP1108790-A2.
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
PF
XX
XX 16-DEC-1999; 99GP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA
XX Nakagawa S, Mizoguchi H, Ando S, Hayaashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
XX N-PSDB; AAH67170.
DR
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX Claim 17; SEQ ID NO: 5705; 246bp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 378 AA;
SQ
Query Match 40.8%; Score 761; DB 22; Length 378;
Best Local Similarity 44.3%; Pred. No. 2.1e-68;
Matches 159; Conservative 64; Mismatches 124; Indels 12; Gaps 3;
QY 9.RRKSTIYGVNPIGCGAPPAVOSMTNRTTDEATVNOIKALERVGADIVRSVPTMDA 68
DB 5.RRTHOIKVGDVAVGGAPISVOSMTTKTHNDGATLQOIALTAAGCDIVRAVACPTDXD 64

QY 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNDMDIDQLEARIRAKSOLD 360
 DB 301 SIIGCVNNGPGEALVSTLGTGNGKSGLYEDGVKRDRLDNDMDIDQLEARIRAKSOLD 360
 QY 361 EARRIDVQVQVEK 372
 DB 361 EARRIDVQVQVEK 372
 RESULT 3
 ABB47958
 ID ABB47958 standard; Protein; 368 AA.
 AC ABB47958;
 XX
 DE 05-FEB-2002 (first entry)
 XX
 XX Listeria monocytogenes protein #662.
 DE Listeria monocytogenes protein #662.
 XX
 KM Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
 KM vitamin B12; bacterial infection; disease.
 XX
 OS Listeria monocytogenes.
 XX
 PN WO200177335-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR01118.
 XX
 PR 11-APR-2000; 2000FR-0004629.
 XX
 PA (INSP) INST PASTEUR.
 XX
 PI Buchrieser C, Frangeul L, Couve E, Rusniok C, Feihl H, Dehoux P,
 PI Dussurget O, Chetoui F, Nedjati H, Glaser P, Kunst F, Cossart P,
 PI Daniels J, Goebel W, Krefz J, Kuhn M, Ng E, Vazquez-Boland JA,
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
 PI Chakraborty T, Domann E, Hain T, Berche P, Chabdit A, Durant L,
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
 PI Madueno E, De Pablo B, Weiland J, Kaerst U, Entian K, Hauf J,
 PI Rose M, Voss H;
 XX
 DR MPI; 2002-010914/01.
 XX
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -
 XX
 PS Claim 6; SEQ ID No 663; 192pp; French.
 XX
 CC The present invention relates to the genome sequence of Listeria
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 368 AA;
 Query Match 46.5%; Score 868; DB 23; Length 368;

Best Local Similarity 49.2%; Pred. No. 2.6e-79;
 Matches 178; Conservative 67; Mismatches 115; Indels 2; Gaps 2;
 QY 10 RKSTR-IYGNVPDYGAPLAVQSMRTTDTVEATVNOIKALERGADIVRSVPTMDA 68
 DB 7 RENTRPVQVGNLTIGSEBELTIGSMTTTTHDVEATVBIHRLSEAGCOIWRVACPDERA 66
 QY AEAFLIKOOVNPPLVADIHFDYRIALKVAEYGVDCILRNPNIGNIEERIIRWVDCARDK 128
 DB 67 ANLSAIKKKIHPLVADIHFDYRIALKAIDACVDYKIRINPNIGNIEERIIRWVDCARDK 126
 QY 129 NIPRIGVAGSLIEKLOEKYGEPTPQALLSEAMRHVDLDRINPDQFVSVKASDVFLA 188
 DB 127 NIPRIGVAGSLIEKTIQKGYPTADGWESALAHIKILEDIDFYDIIVSLKASDVNLA 186
 QY 189 VESYRLAKQIDQPHLGTTEAGARSGAVKASIGILLSEIGTTLVSLAADVEEL 248
 DB 187 IEAYDKASRAFNPLHLGITTESGTOPAGGIKSAAGAILSLIGNTLVSLADVEEL 246
 QY 249 KVGFDLIKSLRISRGINFACPTCSROEFVDYGVNVALEORLEDIITPMDSIICVYN 308
 DB 247 KVAREVLKSPGLSSNAAMLSIPTCGRIIDILRIANEVENYIAKEVEIKVAVGCAYN 306
 QY 309 GPGELVSTLGTGNGKSGLYEDGVKRDRLDNDMDIDQLEARIRAKSOLD EARRIDVQ 368
 DB 307 GPGELVSTLGTGNGKSGLYEDGVKRDRLDNDMDIDQLEARIRAKSOLD EARRIDVQ 368
 QY 369 QV 370
 DB 366 SL 367

RESULT 4
 ABB65322
 ID ABB65322 standard; Protein; 403 AA.
 XX
 AC ABB65322;
 XX
 DT 19-NOV-2002 (first entry)
 XX
 DE Bifidobacterium longum NCC2705 ORF amino acid sequence SEQ ID NO:66.
 XX
 KM Bifidobacterium longum NCC2705; Bifidobacterium; bacterial;
 KM anti-diarrhetic; antibacterial; inhibitor of Salmonella; detection;
 KM identification; lactic acid bacterium; diarrhoea; pathogenic bacteria;
 KM rotavirus; food composition; pharmaceutical composition.
 XX
 OS Bifidobacterium longum.
 XX
 PN EP1227152-A1.
 XX
 PD 31-UTL-2002.
 XX
 PF 30-JAN-2001; 2001EP-0102050.
 XX
 PR 30-JAN-2001; 2001EP-0102050.
 XX
 PA (NEST) SOC PROD NESTLE SA.
 XX
 DR MPI; 2002-668397/72.
 XX
 PT Novel polynucleotide comprising Bifidobacterium genome sequence useful
 PT as a probe or primer for detecting and/or identifying Bifidobacterium
 PT longum in a biological sample -
 XX
 PS Claim 3; SEQ ID 66; 80pp; English.
 CC The present invention describes a polynucleotide (I) comprising a
 CC sequence of a Bifidobacterium genome selected from the nucleotide
 CC sequences given in ABO81842 and ABO81843, or a sequence exhibiting at
 CC least 90% identity or which hybridises with the sequences given in
 CC ABO81842 and ABO81843. Also described is a polynucleotide (II) encoding
 CC a fusion protein, comprising a sequence selected from 1097 sequences

PT increasing isoprenoid content and identifying e.g. antimicrobial
 PT agents, comprises using DNA sequences from bacteria or parasites
 PS Disclosure; Page 15-17; 36pp; German.

XX This invention describes a novel method for incorporating gcpe and yfegB
 CC genes into viruses and cells for increasing isoprenoid content and
 CC identifying e.g. antimicrobial agents, comprises using DNA sequences (I)
 CC from the gcpe or yfegB genes of bacteria or parasites or DNA sequences
 CC (II) which hybridize to the specified genes or encode a protein
 CC with the same biological activity as those encoded by the genes. The
 CC invention also describes (1) plant cells containing (I) or (II); (2)
 CC transformed plant cells, and transgenic plants regenerated from them,
 CC that contain (I) or (II); (3) determining the enzymatic activity of a
 CC gcpe protein; or (4) screening compounds (A) that have antimycotic,
 CC antiparasitic or antiviral activity in humans or animals or antiviral,
 CC antiparasitic, fungicidal or herbicidal activity in plants. (I) and (II)
 CC are used: (i) to increase the isoprenoid levels in viruses and cells;
 CC (ii) for determining the enzymatic activity of gcpe and yfegB proteins;
 CC and (iii) to identify compounds that inhibit activity of gcpe, i.e.
 CC potential antibacterial, antimycotic, antiparasitic or antiviral agents
 CC for use in humans or animals, or antiviral, antiparasitic, fungicidal
 CC or herbicidal agents for agriculture.

XX Sequence 372 AA;

Query Match 100.0%; Score 1866; DB 22; Length 372;
 Best Local Similarity 100.0%; Pred. No. 7.6e-181;
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHNQAPIORRKRSTRIVYGNVPIGDGAPIAVQSMNTTTRTVEATVNOIKALERVGADIVR 60
 DB 1 MHNQAPIORRKRSTRIVYGNVPIGDGAPIAVQSMNTTTRTVEATVNOIKALERVGADIVR 60
 QY 61 VSVPTMDAAEAFKLIRKQVNVPLVADIHFDYRIALKAAYGVDCRLINPNIENGERRIM 120
 DB 61 VSVPTMDAAEAFKLIRKQVNVPLVADIHFDYRIALKAAYGVDCRLINPNIENGERRIM 120
 QY 121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPQALLSMMRHVDLRLNFDQFVSV 180
 DB 121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPQALLSMMRHVDLRLNFDQFVSV 180
 QY 181 KASDVFLAVESYRLAKQIDQPHLGTTEAGARSAGVKSALIGLLSISGIDTLRVSL 240
 DB 181 KASDVFLAVESYRLAKQIDQPHLGTTEAGARSAGVKSALIGLLSISGIDTLRVSL 240
 QY 241 AADPVEIKYGFDLKSLRISRGINFACPTCSRQEFVDVIGTVNALQERLEDIITPMDV 300
 DB 241 AADPVEIKYGFDLKSLRISRGINFACPTCSRQEFVDVIGTVNALQERLEDIITPMDV 300
 QY 301 SIIGCVNNGPGEALVSTLGVGTGKKSGLYEDGVKRDLDNNMDIDQLRARIRAKASQLD 360
 DB 301 SIIGCVNNGPGEALVSTLGVGTGKKSGLYEDGVKRDLDNNMDIDQLRARIRAKASQLD 360
 QY 361 EARRIDVOQVEK 372
 DB 361 EARRIDVOQVEK 372

RESULT 2

AAE19653 standard; Protein; 372 AA.

AAE19653;

31-MAY-2002 (first entry)

Escherichia coli GCPE protein.

gcpe gene; methyl-D-erythritol phosphate; MEP, dimethylallyl diphosphate;
 transgenic plant; isoprenoid compound; tocopherol; isopentenyl diphosphate;
 food; feed source; transfection; single nucleotide polymorphism; SNP;
 oxidative stress tolerance; UV tolerance; transformation; GCPE protein;

KW plant.

OS Escherichia coli.

PN WO200212478-A2.

PD 14-FEB-2002.

PF 06-AUG-2001; 2001WO-US24335.

PR 07-AUG-2000; 2000US-223483P.

(MONS) MONSANTO TECHNOLOGY LLC.

PA Boronat A, Campos N, Rodriguez-concepcion M, Rohmer M, Seeman M;

PI Valentin HE, Venkatesh TV, Venkatesh M,

DR WPI; 2002-227151/28.

PS N-FSDB; AAD31203.

PT gcpe nucleic acid which is an essential gene of the methyl-D-erythritol
 PT phosphate pathway, encoding a fully defined GCPE protein which is
 PT useful for increasing levels of tocopherol substrates in plants
 PS Claim 1; Page 144-145; 155pp; English.

XX The invention relates to gcpe nucleic acid molecule, an essential gene
 CC of methyl-D-erythritol phosphate (MEP) pathway that encodes rice,
 CC Arabidopsis thaliana or Escherichia coli GCPE protein. gcpe is useful
 CC for producing a transgenic plant such as Brassica campestris, B. napus,
 CC canola, castor bean, coconut, cotton, cranberry, linseed, maize, mustard,
 CC oil palm, peanut, rapeseed, rice, safflower, sesame, soybean, sunflower,
 CC or wheat with an increased isoprenoid (tocopherol) compound level. The
 CC expression of GCPE protein in organisms increases the level of
 CC tocopherol substrate such as isopentenyl diphosphate and dimethylallyl
 CC diphosphate biosynthesis. Transgenic organisms overexpressing GCPE
 CC protein can nutritionally enhance food and feed sources. Overexpression
 CC of GCPE protein in transgenic plant may provide tolerance to stresses
 CC e.g., oxidative stress tolerance such as to oxygen or ozone, UV
 CC tolerance, etc. gcpe may be used to obtain nucleic acid molecules from
 CC the same species, and to obtain nucleic acid homologues. gcpe is also
 CC used as or primers. The recombinant vectors are used in plant
 CC transformation or transfection. gcpe an also act as markers capable of
 CC detecting polymorphisms such as single nucleotide polymorphisms (SNPs).
 CC gcpe is also used to determine the level or pattern of expression of
 CC the protein. The present sequence is Escherichia coli GCPE protein.

XX Sequence 372 AA;

Query Match 100.0%; Score 1866; DB 23; Length 372;
 Best Local Similarity 100.0%; Pred. No. 7.6e-181;
 Matches 372; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHNQAPIORRKRSTRIVYGNVPIGDGAPIAVQSMNTTTRTVEATVNOIKALERVGADIVR 60
 DB 1 MHNQAPIORRKRSTRIVYGNVPIGDGAPIAVQSMNTTTRTVEATVNOIKALERVGADIVR 60
 QY 61 VSVPTMDAAEAFKLIRKQVNVPLVADIHFDYRIALKAAYGVDCRLINPNIENGERRIM 120
 DB 61 VSVPTMDAAEAFKLIRKQVNVPLVADIHFDYRIALKAAYGVDCRLINPNIENGERRIM 120
 QY 121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPQALLSMMRHVDLRLNFDQFVSV 180
 DB 121 VVDCARDKNIPIRIGVNAAGSLKDLQEKYGEPTPQALLSMMRHVDLRLNFDQFVSV 180
 QY 181 KASDVFLAVESYRLAKQIDQPHLGTTEAGARSAGVKSALIGLLSISGIDTLRVSL 240
 DB 181 KASDVFLAVESYRLAKQIDQPHLGTTEAGARSAGVKSALIGLLSISGIDTLRVSL 240
 QY 241 AADPVEIKYGFDLKSLRISRGINFACPTCSRQEFVDVIGTVNALQERLEDIITPMDV 300
 DB 241 AADPVEIKYGFDLKSLRISRGINFACPTCSRQEFVDVIGTVNALQERLEDIITPMDV 300

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2003, 15:54:02 ; Search time 51 Seconds
(without alignments)
1157.770 Million cell updates/sec

Title: US-09-921-992-50
Perfect score: 1866
Sequence: 1 MHNQAPIQRKRSTRIVGVN.....RAKASQLEARRIDVOQVEK 372

Scoring table: BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03:*

1: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1866	100.0	372	22	AB45692
2	1866	100.0	372	23	AB45692
3	868	46.5	368	23	AB47958
4	814.5	43.6	403	23	AB65322
5	794	42.6	359	19	AAW8298
6	785.5	42.1	415	22	AAU56212
7	761	40.8	378	22	AAU9151
8	751.5	40.3	387	22	AAU81205
9	751.5	40.3	387	24	AB57500

10	534	28.6	421	24	ABP78786
11	490	26.3	611	20	AAJ37164
12	470.5	25.2	621	20	AAJ34971
13	462	24.8	740	23	AAE19650
14	460	24.7	679	22	AB45693
15	446	23.9	752	21	AAW0883
16	445.5	23.9	603	23	AAE19652
17	445.5	23.9	686	23	AAE19651
18	428.5	23.0	169	21	AAE1371
19	428.5	23.0	169	23	ABP32470
20	398	21.3	144	23	ABU52039
21	242	13.0	776	22	ABG25879
22	119	6.4	113	21	AAU08755
23	119	6.4	113	21	AAU08755
24	113	6.1	489	22	AAU96357
25	111.5	6.0	578	20	AAU89273
26	110	5.9	108	21	AAU08756
27	110	5.9	108	21	AAU08756
28	109	5.8	256	22	AAU53170
29	106.5	5.7	284	23	ABU53922
30	105.5	5.7	578	19	AAU82635
31	105.5	5.7	578	21	AAU78566
32	105.5	5.7	578	23	ABG93408
33	105.5	5.7	578	23	AAU73205
34	105	5.6	545	20	AAU26187
35	104	5.6	606	23	ABU48094
36	102.5	5.5	542	23	ABU53701
37	102	5.5	546	23	ABP26321
38	102	5.5	546	23	ABP26321
39	101	5.4	538	23	ABP28474
40	101	5.4	523	22	ABU49213
41	100	5.4	343	19	AAU71467
42	100	5.4	343	21	AAU99880
43	100	5.4	502	22	AAU99843
44	99	5.3	541	21	AAU91970
45	99	5.3	268	22	AAU34139

ALIGNMENTS

RESULT 1	
AA45692	AA45692 standard; Protein; 372 AA.
XX	
AC	AA45692;
XX	
DT	15-MAR-2001 (first entry)
XX	
DE	E. coli gcpe protein.
XX	
KW	Isoprenoid; gcpe; yfB; antimicrobial; transgenic plant; agriculture;
KW	antimicrobial; antiparasitic; antiviral; fungicidal; herbicidal.
XX	
OS	Escherichia coli.
XX	
PN	MO200072022-A1.
PD	30-NOV-2000.
XX	
PF	20-MAY-2000; 2000WO-EP04592.
XX	
PR	21-MAY-1999; 99DE-1023567.
PR	21-MAY-1999; 99DE-1023568.
XX	
PA	(JOMA/) JOMAA H.
XX	
PI	Jomaa H;
XX	
DR	WPI; 2001-025196/03.
XX	
PT	N-PSDB; AAC82653.
	Incorporating gcpe and yfB genes into viruses and cells, for